Psychometrika

CONTENTS

VOLUME TWENTY-THREE SEPTEMBER 1958 NUMBER 3

PSYCHOMETRIKA—VOL. 23, NO. 3
SEPTEMBER, 1958

THE VARIMAX CRITERION FOR ANALYTIC ROTATION IN FACTOR ANALYSIS*

HENRY F. KAISER
UNIVERSITY OF ILLINOIS

An analytic criterion for rotation is defined. The scientific advantage of analytic criteria over subjective (graphical) rotational procedures is discussed. Carroll's criterion and the quartimax criterion are briefly reviewed; the varimax criterion is outlined in detail and contrasted both logically and numerically with the quartimax criterion. It is shown that the *normal* varimax solution probably coincides closely to the application of the principle of simple structure. However, it is proposed that the ultimate criterion of a rotational procedure is factorial invariance, not simple structure—although the two notions appear to be highly related. The normal varimax criterion is shown to be a two-dimensional generalization of the classic Spearman case, i.e., it shows perfect factorial invariance for two pure clusters. An example is given of the invariance of a normal varimax solution for more than two factors. The oblique normal varimax criterion is stated. A computational outline for the orthogonal normal varimax is appended.

In factor analysis, an analytic criterion for rotation is defined as one that imposes mathematical conditions beyond the fundamental factor theorem, such that a factor matrix is uniquely determined. Historically, the first such criterion was Thurstone's treatment of the principal axes problem [10]: from any arbitrary factor matrix he suggested rotating under the criterion that each factor successively accounts for the maximum variance. But principal axes have seldom been accepted as psychologically very interesting ([9], p. 139). The rotation problem for psychologically meaningful factors is usually handled judgmentally. Scientifically, however, this procedure is not very satisfactory: the ad hoc quality of subjective rotation makes uniquely determined factors impossible; only factors that are subject to the uncertainties and controversies besetting any a posteriori reasoning can be defined. In contrast, an analytic criterion for rotation would allow factor analysis to become a straightforward methodology stripped of its subjectivity and a proper tool for scientific inquiry.

The Quartimax Criterion

The first analytic criterion for determining psychologically interpretable factors was presented in 1953 by Carroll [1]. In an attempt to provide a

*Part of the material in this paper is from the writer's Ph.D. thesis. I am indebted to my committee, Professors F. T. Tyler, R. C. Tryon, and H. D. Carter, chairman, for many helpful suggestions and criticisms. Dr. John Caffrey suggested the name *varimax*, and wrote the original IBM 602A computer program for this criterion.

I am also indebted to the staff of the University of California Computer Center for help in programming the procedures described in the paper for their IBM 701 electronic computer. Since their installation is partially supported by a grant from the National Science Foundation, the assistance of this agency is acknowledged.

mathematical explication of Thurstone's simple structure, he suggested that for a given factor matrix,

$$f = \sum_{s<t} \sum_j a_{js}^2 a_{jt}^2 \tag{1}$$

should be a minimum, where $j = 1, 2, \cdots, n$ are tests, $s, t = 1, 2, \cdots, r$ are factors, and a_{js} is the factor loading of the jth test on the sth factor. It appears that Carroll was motivated in writing (1) primarily by a close inspection of Thurstone's five formal rules for simple structure ([12], p. 335), particularly the requirement that a large loading for one factor be opposite a small loading for another factor.

In his original paper, Carroll provided two numerical examples of the application of his method. Without the restriction of orthogonality, these illustrations gave somewhat equivocal results—while the application of (1) appears to bring one close to the desired simple structure, the criterion has an obvious bias in being too strongly influenced by factorially complex tests.

In the light of later developments, Carroll's criterion should probably be relegated to the limbo of "near misses"; however, this does not detract from the fact that it was the first attempt to break away from an inflexible devotion to Thurstone's ambiguous, arbitrary, and mathematically unmanageable qualitative rules for his intuitively compelling notion of simple structure.

Almost simultaneously with Carroll's development, Neuhaus and Wrigley [7], Saunders [8], and Ferguson [2] proposed what is usually called the quartimax method for orthogonal simple structure. Neuhaus and Wrigley suggest that a most easily interpretable factor matrix, in the simple structure sense, may be found when the variance of all nr squared loadings of the factor matrix is a maximum, i.e.,

$$q_1 = [nr \sum_j \sum_s (a_{js}^2)^2 - (\sum_j \sum_s a_{js}^2)^2]/n^2r^2 = \text{maximum}. \tag{2}$$

Saunders' approach requires that the kurtosis (fourth moment over second moment squared) of all loadings and their reflections be a maximum,

$$q_2 = nr \sum_j \sum_s a_{js}^4 / (\sum_j \sum_s a_{js}^2)^2 = \text{maximum}. \tag{3}$$

While Ferguson, basing his rationale on certain parallels with information theory, calls simply for

$$q_3 = \sum_j \sum_s a_{js}^4 = \text{maximum}. \tag{4}$$

All these investigators are concerned with attaining a factor matrix with a maximum tendency to have both small and large loadings. While less obviously related to Thurstone's rules than Carroll's criterion, the emphasis on small loadings coincides with Thurstone's requirements of

zero loadings. For orthogonal factors, criteria (2), (3), and (4) are equivalent because of the invariance of the sum of the communalities, $\sum_j \sum_s a_{js}^2$. (This term, as well as other constants, disappear when differentiated in the calculus problem involved in finding the required critical point.)

Indeed, it turns out that they are also equivalent to Carroll's criterion in the orthogonal case. Minimizing (1) is equivalent to maximizing (4) since the squared communality of a test is

$$\text{constant} = (\sum_s a_{js}^2)^2 = \sum_s a_{js}^4 + 2\sum_{s<t} a_{js}^2 a_{jt}^2,$$

and the sum of squared communalities over all tests is

$$\text{constant} = \sum_j \sum_s a_{js}^4 + 2\sum_j \sum_{s<t} a_{js}^2 a_{jt}^2$$

$$= q_3 + 2f.$$

Thus, since the quartimax criterion plus twice Carroll's criterion is a constant, maximizing q_3 is equivalent to minimizing f.

Neuhaus and Wrigley realized that none of these criteria can be realistically applied without the aid of an electronic computer—the calculations involved are too extensive for a desk calculator or punched card mechanical computers. Consequently, they programmed the quartimax method for the Illiac* and provided a rather extensive numerical investigation of the empirical properties of the quartimax method.

Their results were perhaps more encouraging than Carroll's. Under the restriction of orthogonality Carroll's criterion (or the equivalent quartimax method) does not show nearly so obvious a bias as does Carroll's criterion when the restriction of orthogonality is removed. However, as an explication of orthogonal simple structure, the quartimax method does have a systematic bias which will be more fully examined in the next section.

The Varimax Criterion

From the outset, the above methods consider all nr loadings simultaneously. In every case, however, these criteria may be applied separately to each *row* of the factor matrix and summed over rows for the final criterion because of the invariance of the communalities. For example, Neuhaus and Wrigley could have defined the *simplicity*, say, of the factorial composition of the jth test as the variance of the squared loadings for this test,

$$q_j^* = [r \sum_s (a_{js}^2)^2 - (\sum_s a_{js}^2)^2]/r^2. \tag{5}$$

*The Illiac is the University of Illinois electronic computer. Subsequently, the quartimax criterion has been programmed for the CRC–102A (Neuhaus), and the IBM 701 (Kaiser). The varimax criterion described in the next two sections has been programmed for SWAC at UCLA (Comrey), the IBM 701 (Kaiser), Illiac (Dickman), and the IBM 650 (Vandenberg).

To obtain the total criterion for the entire factor matrix, (5) could then be summed over all tests to give

$$q^* = \sum_i \{[r \sum_s (a_{is}^2)^2 - (\sum_s a_{is}^2)^2]/r^2\}. \tag{6}$$

Maximizing q^* is equivalent to maximizing q_3, again because constant terms vanish when differentiated.

Equation (6) perhaps provides some insight into the quartimax criterion—its aim is to simplify the description of each row, or test, of the factor matrix. It is unconcerned with simplifying the columns, or factors, of the factor matrix (probably the most fundamental of all requirements for simple structure). The implication of this is that the quartimax criterion will often give a general factor. Under requirement (5) there is no reason why a large loading for each test may not occur on the same factor. In practice, this tendency for the quartimax criterion to yield a general factor is most pronounced when the unrotated factor matrix has a pronounced general factor.

From the simple structure viewpoint, an immediate modification of the quartimax criterion is apparent. Let us define the simplicity of a factor as the variance of its squared loadings,

$$v_s^* = [n \sum_j (a_{js}^2)^2 - (\sum_j a_{js}^2)^2]/n^2. \tag{7}$$

And for the criterion for all factors, define the maximum simplicity of a factor matrix as the maximization of

$$v^* = \sum_s v_s^* = \sum_s \{[n \sum_j (a_{js}^2)^2 - (\sum_j a_{js}^2)^2]/n^2\}, \tag{8}$$

the variance of squared loadings by columns rather than by rows.

Since a factor is a vector of correlation coefficients, the most interpretable factor is one based upon correlation coefficients which are maximally interpretable. Those correlations which satisfy this condition are patently obvious: correlations of $\pm$ 1, which indicate a functional relationship, and correlations of zero, which indicate no linear relationship. On the other hand, middle-sized correlations are the most difficult to understand. Thus, it is seen why v_s^* in (7) could be maximized for the maximum interpretability or simplicity of a factor, and more generally, why the interpretability of an entire factor matrix could be considered best when (8) is a maximum.

Criterion (8) is the original *raw* varimax criterion [4]. In the original proposal of this criterion, it was shown to be mathematically equivalent, in the orthogonal case, to minimizing

$$c^* = \sum_{s<t} \{[n \sum_j a_{js}^2 a_{jt}^2 - (\sum_j a_{js}^2)(\sum_j a_{jt}^2)]/n^2\}, \tag{9}$$

i.e., minimizing the covariance of pairs of columns of squared loadings and

summing over all possible pairs of columns for the criterion. Criterion (9) then bears the analogous relationship to Carroll's criterion (1) that the varimax criterion (8) does to the quartimax criterion (6).

Some distinctions between quartimax and varimax orthogonal solutions can be illustrated numerically. In Table 1 solutions for Thurstone's eleven-variable box problem ([12], pp. 373–375) are given. It will be noted that the quartimax solution [7] could hardly be called a simple structure. There is a large general factor, and the second factor seems only vaguely concerned

Table 1

Thurstone's 11-Variable Box Problem[a]

	Subjective (oblique)			Quartimax			Raw Varimax		
Test	X	Y	Z	X	Y	Z	X	Y	Z
x	90	05	00	68	65	05	91	19	16
y	04	88	01	83	-47	00	05	93	25
z	03	05	79	42	-08	79	11	17	88
xy	62	63	-06	99	11	-04	64	74	20
yz	-05	54	57	71	-40	56	02	65	75
x^2y	82	37	-01	92	41	03	84	51	22
xy^2	35	76	02	96	-18	03	37	86	28
$2x + 2y$	53	71	-09	100	00	-07	54	82	18
$(x^2 + y^2)^{\frac{1}{2}}$	52	71	-08	99	-01	-07	53	81	18
$(x^2 + z^2)^{\frac{1}{2}}$	52	-07	65	59	38	68	60	09	77
xyz	42	43	43	88	04	45	48	58	65

[a]Decimal points omitted.

with dimensions of boxes. On the other hand, the raw varimax solution closely parallels Thurstone's original subjective solution, given the restriction of orthogonality.

In Table 2 are solutions for Holzinger and Harman's 24 psychological tests ([3], pp. 229–233). Both the quartimax [7] and the raw varimax methods seem to duplicate the subjectively rotated simple structure patterns. But the respective variance contributions of the factors are perhaps more interesting. It is seen that the dispersion of the $\sum_i a_{is}^2$ for the subjective solution is less than the corresponding figures for the two analytic methods. In other words, Holzinger and Harman have made the factors a little more level or

Table 2

Holzinger and Harman's Twenty-four Psychological Tests[a]

	Subjective				Quartimax				Raw Varimax				Normal Varimax			
Test	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D
1	10	32	62	20	37	19	60	07	24	20	65	13	14	19	67	17
2	07	15	41	13	24	07	38	04	17	07	42	08	10	07	43	10
3	10	12	53	13	31	01	48	01	22	02	52	06	15	02	54	08
4	15	18	53	12	36	07	46	-01	27	08	52	04	20	09	54	07
5	75	15	26	15	81	14	-02	-04	78	21	12	06	75	21	22	13
6	72	05	28	25	81	03	00	06	78	10	13	14	75	10	23	21
7	81	08	27	11	85	07	-04	-10	84	15	10	00	82	16	21	08
8	54	26	38	14	66	20	20	-04	60	25	31	05	54	26	38	12
9	76	-04	29	30	86	-06	-02	10	84	01	12	19	80	01	22	25
10	28	66	-19	14	23	70	-12	11	17	71	-08	17	15	70	-06	24
11	27	61	-04	29	31	62	01	23	22	63	06	29	17	60	08	36
12	13	72	09	03	16	69	19	-01	06	70	23	04	02	69	23	11
13	24	63	31	02	35	57	32	-08	24	59	39	-01	18	59	41	06
14	23	19	-02	48	32	19	-03	42	26	20	01	46	22	16	04	50
15	11	14	08	50	25	11	10	45	17	11	13	48	12	07	14	50
16	05	22	34	45	29	13	37	37	17	13	41	41	08	10	41	43
17	15	24	-03	62	28	24	02	57	20	23	05	61	14	18	06	64
18	01	39	20	52	22	32	30	47	08	31	32	51	00	26	32	54
19	12	22	18	39	28	18	19	32	19	18	22	36	13	15	24	39
20	31	18	46	29	52	09	35	14	42	12	43	21	35	11	47	25
21	17	46	33	24	35	38	35	14	23	40	40	20	15	38	42	26
22	31	12	40	40	53	04	30	26	44	06	37	32	36	04	41	36
23	31	29	54	25	55	19	44	09	43	21	52	16	35	21	57	22
24	39	46	14	31	49	43	10	20	40	46	18	27	34	44	22	34
$\sum_j a_{js}^2$	343	292	268	236	559	242	196	142	431	260	264	186	350	244	308	236

[a]Decimal points omitted.

even in their contribution to variance than the analytic criteria. Of the two analytic criteria, the raw varimax solution has given a solution which is closer in this respect to Holzinger and Harman's. It is also noteworthy that as a result of these differences the large loadings of the factors with the larger variance contributions for the analytic methods are larger than the large loadings for the smaller factors, and similarly, the small loadings for the larger factors are larger than the small loadings for the smaller factors. Holzinger and Harman's subjective solution does not show this systematic bias; their solution gives a more equitable patterning of factor loadings.

How this bias may be removed is indicated in the next section. This leads to a revision of the varimax criterion, which appears to have more important characteristics than merely satisfying the rules of simple structure.

Factorial Invariance: Normal Varimax

It seems reasonable to attribute the systematic bias seen in both the quartimax and varimax solutions of the Holzinger-Harman data and other examples [4] to the divergent weights which implicitly are attached to the tests by their communalities. When one deals with fourth-power functions

of factor loadings, a test with communality 0.6, for example, would tend to influence the rotations four times as much as a test whose communality was 0.3. Thus, while the most obvious weights have been applied to the tests, namely the square roots of their communalities, after the fact it seems that there is probably a better set of weights—weights which would tend to equalize to a greater extent the relative influence of each test during rotation.

There seems no rational basis for choosing among different weighting schemes. Let us then make the agnostic confession of ignorance which pervades any form of correlational analysis. For the purposes of rotation, weight the tests equally, in the sense that the lengths of the common parts of the test vectors have equal length. (The author is indebted to Dr. D. R. Saunders for this suggestion.) The varimax criterion could then be rewritten as

$$v = \sum_{s} \{[n \sum_{j} (a_{js}^2/h_j^2)^2 - [\sum_{j} (a_{js}^2/h_j^2)]^2]/n^2\}, \tag{10}$$

where h_j^2 is the communality of the jth test. In contrast to (7) and (8), where the variance of the squared correlations of the tests with a factor is maximized, the variance of the squared correlations of the common parts of the tests (the reflections of the tests onto the common-factor space) with a factor is now being maximized. [Note from (10) that we are not advocating a permanent weighting of the tests by a weight inversely as the square root of their communalities. During rotation this weighting extends the common part of each test vector to unit length, but after rotation each of these vectors is shortened to its proper length by reweighting directly as the square root of the test's communality.]

As will be seen in Table 2, under this modification the varimax criterion (the *normal* varimax, since rotation is with respect to normalized common parts of tests) has effectively removed the small but disturbing bias in the raw varimax solution of Holzinger and Harman's example. It also has been shown in a number of other examples [6] that the normal varimax does not seem to deviate systematically from what may be considered the best orthogonal simple structure.*

Thus far, however, merely a numerical-intuitive basis for a weighting procedure which leads to "prettier" results has been provided. Such a basis is quite unsatisfactory theoretically. Indeed, this sort of ad hoc thinking could conceivably lead to a different set of judgmentally determined weights for any particular example—a situation as scientifically reprehensible as the subjective graphical methods.

There is a more fundamental rationale for attempting to establish the normal varimax criterion (10) as a mathematical definition for the rotation

*Professor Andrew Comrey has apparently reached the same conclusion in an extensive application of the normal varimax criterion to interitem correlation matrices of the MMPI (personal communication). A further example, available from the writer, is the normal varimax solution of Thurstone's classic PMA study[11] (dittoed).

problem. Consider the situation illustrated in Fig. 1. There are two clusters of tests, each of which is pure in the sense that the reflections of the test vectors of the cluster onto the two-dimensional, common-factor space are collinear. (While these clusters are drawn less than 90° apart, the following argument is perfectly general.)

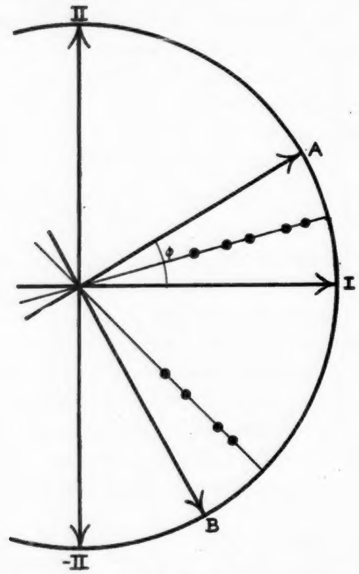

FIGURE 1

Case for which a normal varimax solution is invariant under changes in the composition of the test battery.

It is shown below that the angle of rotation in a plane which maximizes (10) is

$$\phi = \tfrac{1}{4} \arctan \frac{2[n \sum_j u_j v_j - \sum_j u_j \sum_j v_j]}{n \sum_j (u_j^2 - v_j^2) - [(\sum_j u_j)^2 - (\sum_j v_j)^2]}, \tag{11}$$

where

$$u_j = (a_{j1}/h_j)^2 - (a_{j2}/h_j)^2,$$

and

$$v_j = 2(a_{j1}/h_j)(a_{j2}/h_j).$$

Let n_A ($n_A \geq 1$) be the number of tests in the first cluster and n_B ($n_B \geq 1$) be the number of tests in the second cluster ($n = n_A + n_B$). It is readily apparent that all tests of the first cluster have the same values for u_j and v_j.

Let these values be u_A and v_A. Similarly let the values for the second cluster be u_B and v_B. In this case (11) reduces to

$$\phi = \tfrac{1}{4} \arctan \frac{2n_A n_B(u_A v_A + u_B v_B - u_A v_B - u_B v_A)}{n_A n_B(u_A^2 + u_B^2 - v_A^2 - v_B^2 - 2u_A u_B + 2v_A v_B)} . \tag{12}$$

A most important result is shown in (12). The $n_A n_B$ term may be cancelled, indicating that the angle of rotation does not depend on the number of tests in each cluster, i.e., *for the case illustrated in Fig.* 1, *the normal varimax solution is invariant under changes in the composition of the test battery.*

This invariance property would seem to be of greater significance than the numerical tendencies of the normal varimax solution to define mathematically the doctrine of simple structure. Although factor analysis seems to have many purposes, fundamentally it is addressed to the following problem. Given an (infinite) domain of psychological content, infer the internal structure of this domain on the basis of a sample of n tests drawn from the domain. The possibility of success in such inferences is obviously dependent upon the extent which a factor derived from a particular battery or sample of tests approximates the corresponding unobservable factor in the infinite domain. If a factor is invariant under changing samples of tests, i.e., shows factorial invariance ([12], pp. 360–361), there is evidence that inferences regarding domain factors are correct.

The normal varimax solution, according to the above result, allows such inferences; regardless of the sampling of tests, for the problem shown in Fig. 1 it is possible to infer precisely the domain normal varimax factors. This is not true for either the quartimax or raw varimax solutions since the angle of rotation is a function of n_A and n_B.

Note that domain normal varimax factors are not said to be more *meaningful* than domain factors according to some different criterion; it is suggested that observed normal varimax factors will have a greater likelihood of portraying the corresponding domain factors.

Although one often gets the impression that simple structure is the ultimate criterion of a rotational procedure, it is suggested here that the ultimate criterion is factorial invariance. The normal varimax solution was originally devised solely for the purpose of satisfying the simple structure criteria. But the fact that it shows mathematically this sort of invariance suggests that Thurstone's reasoning was basically directed toward factorial invariance. The principle of simple structure may probably be considered incidental to the more fundamental concept of factorial invariance. This viewpoint renders meaningless the arguments concerning "psychological reality" of general factors, bipolar factors, simple structure factors, etc.

Admittedly, the result (12) is for a special case. The correlations among the variables within each of the two pure clusters must form a perfect Spearman matrix, and the reduced correlation matrix as a whole must be

of rank two. Normal varimax does however give invariant results for two such Spearman clusters simultaneously, and consequently the normal varimax criterion is a two-dimensional generalization of the classic Spearman case. Obviously, the next step would be a generalization along the same lines to the r-dimensional case; thus far, however, work on this problem suggests virtually insuperable mathematical difficulties.

To investigate numerically the tendency of the criterion to give factorially invariant solutions for r larger than two, again consider Holzinger and Harman's empirical data. Taking their centroid loadings, the first five tests were rotated then the first six, etc., systematically until the analysis of all 24 tests was reached, as in Table 2. The results of this application of the normal varimax criterion are given in Table 3. There, the normal varimax loadings for the four factors as a function of the changing number of tests are given.

Note that factors A and C are essentially invariant from the outset; the loading changes, while somewhat systematic, are negligible—24 appears to be essentially infinite. On the other hand, factors B and D show more movement before they become stable. The reason for this is readily apparent. For both B and D, this movement occurs only while they are underdetermined, i.e., only while they contain no appreciable loadings. However, once they pick up a test or two with high loadings, their ambiguous definition abruptly stops, and they settle down to exhibit a degree of invariance similar to the

Table 3A

Normal Varimax Loading Changes for Holzinger and Harman's Factor A (n = 5, 6, ..., 24)[a]

	n																			
Test	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1	19	19	18	19	18	16	16	16	16	16	15	16	15	16	16	15	15	15	14	14
2	12	12	12	13	12	12	12	12	12	11	11	11	11	12	11	11	11	11	10	10
3	15	16	15	16	16	17	17	17	17	16	16	16	16	17	17	16	16	16	15	15
4	21	21	21	22	21	21	21	21	21	21	21	21	21	21	21	20	20	20	20	20
5	79	78	78	79	78	76	75	76	76	76	75	75	75	76	76	75	75	75	75	75
6		78	77	77	78	77	76	77	77	76	75	75	75	75	75	75	75	75	75	75
7			83	84	83	82	82	82	82	82	82	82	82	83	82	82	82	82	82	82
8				59	58	55	55	55	55	55	55	55	55	55	55	55	55	55	54	54
9					82	83	82	83	83	81	80	80	80	80	80	80	80	80	80	80
10						18	15	17	17	16	16	16	15	16	16	16	16	15	16	15
11							19	21	21	19	18	18	18	18	18	18	18	18	18	17
12								03	02	03	03	03	03	03	03	03	03	03	02	02
13									18	19	19	20	20	20	20	19	19	19	18	18
14										23	21	22	21	22	22	22	22	22	22	22
15											11	12	11	12	12	12	12	12	12	12
16												09	09	10	09	09	09	09	09	08
17													13	14	14	14	14	14	14	14
18														01	01	00	00	00	00	00
19															13	13	13	13	13	13
20																35	35	35	35	35
21																	16	16	16	15
22																		36	36	36
23																			35	35
24																				34

Table 3B

Normal Varimax Loading Changes for Holzinger and Harman's Factor B $(n = 5, 6, \ldots, 24)$[a]

	n																			
Test	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1	10	10	10	10	07	12	16	19	21	20	20	19	19	19	18	19	19	19	19	19
2	01	01	01	01	02	03	05	07	08	08	07	07	07	07	06	07	07	07	07	07
3	-07	-07	-07	-06	-04	-04	-01	01	02	03	02	02	02	02	02	02	02	02	02	02
4	-02	-03	-02	-01	-04	02	05	07	09	09	09	09	09	08	08	08	09	09	08	09
5	-00	03	05	04	-00	19	22	21	22	22	21	21	20	20	20	20	20	21	20	21
6		-03	-02	-05	03	10	14	12	13	11	10	10	09	08	08	08	09	09	09	10
7			-03	-03	-08	12	16	15	16	16	16	16	15	15	15	15	15	16	16	16
8				10	02	21	24	25	26	26	26	26	25	25	25	25	25	25	25	26
9					03	03	08	05	05	03	02	01	00	-00	-01	-01	-00	00	-00	01
10						73	74	74	74	72	71	71	70	69	69	69	69	70	70	70
11							68	67	67	64	62	62	61	60	59	59	60	60	60	60
12								67	68	69	69	69	69	69	68	69	69	69	69	69
13									57	59	59	59	59	59	59	59	59	59	59	59
14										22	19	19	16	15	15	14	15	15	15	16
15											10	10	07	06	06	06	06	06	06	07
16												12	10	09	08	08	09	09	09	10
17													19	17	17	17	17	17	17	18
18														26	25	25	26	26	26	26
19															14	14	14	15	15	15
20																09	10	10	10	11
21																	37	38	37	38
22																		04	04	04
23																			20	21
24																				44

[a]Decimal points omitted.

Table 3C

Normal Varimax Loading Changes for Holzinger and Harman's Factor C $(n = 5, 6, \ldots, 24)$[a]

	n																			
Test	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1	70	70	70	70	70	70	70	69	68	68	68	67	67	67	66	67	67	67	67	67
2	45	44	45	44	44	45	44	44	44	44	44	43	43	43	43	43	43	43	43	43
3	54	54	54	54	54	54	54	54	54	54	54	54	54	54	54	54	54	54	54	54
4	55	55	55	55	55	55	55	55	54	54	54	54	54	54	54	54	54	54	55	54
5	23	23	24	22	23	23	23	21	21	21	21	20	20	20	19	20	20	20	22	22
6		24	25	24	24	24	24	23	23	23	22	21	21	21	20	21	21	21	23	23
7			21	20	21	21	21	20	19	19	19	19	19	18	18	19	19	19	21	21
8				39	40	40	39	38	37	37	37	37	37	36	36	37	37	37	38	38
9					23	23	23	22	22	22	21	20	21	20	19	20	20	20	22	22
10						04	01	-03	-04	-05	-06	-06	-06	-07	-07	-07	-07	-07	-06	-06
11							16	13	11	11	09	09	09	08	07	07	07	07	08	08
12								25	23	23	23	23	23	23	23	23	23	22	23	23
13									41	41	41	41	41	41	41	41	41	41	41	41
14										08	05	04	05	03	02	03	03	03	04	04
15											15	14	15	13	12	13	13	13	14	14
16												42	42	41	40	41	41	41	41	41
17													08	06	04	05	06	06	06	06
18														32	31	31	31	31	32	32
19															22	23	23	23	24	24
20																47	47	47	48	47
21																	42	41	42	42
22																		41	41	41
23																			57	57
24																				22

Table 3D

Normal Varimax Loading Changes for Holzinger and Harman's Factor D (n = 5, 6, ..., 24)[a]

	n																			
Test	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1	01	-00	00	01	-07	01	01	01	02	08	13	14	15	17	19	18	18	17	17	17
2	01	01	02	01	00	02	02	02	01	05	08	09	09	10	11	11	10	10	10	10
3	-00	01	01	-00	05	-00	01	-00	-01	02	06	07	07	08	09	09	08	08	08	08
4	-03	-03	-03	-04	-01	-04	-03	-04	-04	01	05	06	06	07	09	08	08	08	07	07
5	-00	-06	-03	-02	-09	-06	-06	-05	-05	07	12	12	14	14	15	14	14	14	14	13
6		06	09	09	05	05	06	06	05	17	21	21	22	21	22	22	22	22	22	21
7			-06	-06	-06	-10	-10	-09	-10	03	07	07	09	08	09	09	09	08	08	08
8				-04	-14	-07	-08	-07	-07	04	10	10	12	12	14	13	13	13	12	12
9					15	11	12	12	11	22	26	26	27	25	26	26	26	26	25	25
10						-00	-07	-00	02	15	19	19	23	24	25	25	25	24	24	24
11							07	12	15	27	31	32	35	36	37	37	37	37	37	36
12								-14	-11	-00	05	05	09	11	13	12	12	11	11	11
13									-17	-05	00	01	04	07	08	08	07	07	07	06
14										47	49	49	50	50	50	50	50	50	50	50
15											49	49	50	50	50	50	50	50	50	50
16												42	42	44	45	44	44	44	44	43
17													64	64	64	64	64	64	64	64
18														54	55	55	54	54	54	54
19															40	40	40	39	39	39
20																26	26	26	25	25
21																	27	26	26	26
22																		36	36	36
23																			22	22
24																				34

[a]Decimal points omitted.

other two factors, which had high loadings from the beginning. For $n = 24$, there appear to be good approximations to the domain normal varimax factors.

The Oblique Case

If the restriction of orthogonality is relaxed, it is impossible to apply directly the quartimax criterion (4) or the normal varimax criterion (10). This is because interfactor relationships are not considered when the criteria are in this form, and when applied all factors will collapse into the same factor—that one factor which best meets the criterion. However, Carroll's version of the quartimax criterion explicitly considers interfactor relationships and an oblique solution is attainable. As suggested by (9), if

$$c = \sum_{s<t} \{[n \sum_i (a_{js}^2/h_j^2)(a_{jt}^2/h_j^2) - (\sum_i a_{js}^2/h_j^2)(\sum_i a_{jt}^2/h_j^2)]/n^2\}, \tag{13}$$

it may be shown that in the orthogonal case $v = -2c$. This alternative form of the normal varimax may then be used to obtain oblique factors. The mathematical problem of minimizing (13) is exactly analogous to Kaiser's [5] treatment for Carroll's criterion. Computationally, the (iterative) solution involves finding the latent vector associated with the smallest latent root of a constantly changing symmetric matrix of order r.

Computational Appendix

To compute an orthogonal normal varimax solution, the following procedure is suggested. The first step is to normalize the rows of the arbitrary reference factor matrix (e.g., principal axes or centroids) by dividing each element by h_i. Rotation to the direction of the normal varimax factors may then be carried out with respect to these normalized loadings.

The criterion (10) will be applied to two factors at a time. For this purpose, the following notation for an orthogonal rotation is convenient.

$$\begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \\ \vdots & \vdots \\ x_n & y_n \end{bmatrix} \cdot \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix} = \begin{bmatrix} X_1 & Y_1 \\ X_2 & Y_2 \\ \vdots & \vdots \\ X_n & Y_n \end{bmatrix},$$

where x_i and y_i, the present normalized loadings, are constants, and X_i and Y_i, the desired normalized loadings, are functions of ϕ, the angle of rotation.

It is immediately seen that

$$X_i = x_i \cos\phi + y_i \sin\phi, \tag{14}$$

$$Y_i = -x_i \sin\phi + y_i \cos\phi. \tag{15}$$

Thus,

$$dX_i/d\phi = Y_i, \tag{16}$$

$$dY_i/d\phi = -X_i. \tag{17}$$

According to (10), in this plane,

$$n^2 v_{xy} = n \sum (X^2)^2 - (\sum X^2)^2 + n \sum (Y^2)^2 - (\sum Y^2)^2 \tag{18}$$

should be a maximum. Differentiating (18) with respect to ϕ, using (16) and (17), and setting the derivative equal to zero,

$$n \sum XY(X^2 - Y^2) - \sum XY \sum (X^2 - Y^2) = 0. \tag{19}$$

To solve (19) for ϕ in terms of x_i and y_i, substitute the values of X_i and Y_i from (14) and (15), consult a table of trigonometric identities, and, after a good deal of algebraic manipulation,

$$\phi = \tfrac{1}{4} \arctan \frac{2[n \sum (x^2 - y^2)(2xy) - \sum (x^2 - y^2) \sum (2xy)]}{n\{\sum [(x^2 - y^2)^2 - (2xy)^2]\} - \{[\sum (x^2 - y^2)]^2 - [\sum (2xy)]^2\}}. \tag{20}$$

If $u_i = x_i^2 - y_i^2$ and $v_i = 2x_i y_i$, (20) reduces to the form (11) above.

Of course, (11) or (20) is only a necessary condition for a maximum. By taking the second derivative of (18) sufficient conditions for a maximum

may be found. These are summarized below.

		sign of numerator +	sign of numerator −
sign of denominator	+	0° to $+22\frac{1}{2}°$	0° to $-22\frac{1}{2}°$
	−	$+22\frac{1}{2}°$ to $+45°$	$-22\frac{1}{2}°$ to $-45°$

The sign of numerator and denominator refer to the right-hand member of (20); the values in the cells refer to ϕ.

These single-plane rotations are made on factors 1 with 2, 1 with 3, $\cdots$, 1 with r, 2 with 3, $\cdots$, 2 with r, $\cdots$, $(r - 1)$ with r, 1 with 2, $\cdots$ iteratively until $r(r - 1)/2$ successive rotations of $\phi = 0$ are obtained, i.e., until the process converges. (It was shown [6] that v in (10) cannot be greater than $(r - 1)/r$, and since each successive application of (20) can result only in a non-decrease of v, this iterative procedure must converge.) After convergence, each normalized test vector is restored to its proper length by multiplying by h_i.

Since this article was accepted for publication, the author has prepared a detailed outline for coding an electronic computer program for the varimax criterion. This (dittoed) paper is available from the writer.

REFERENCES

[1] Carroll, J. B. An analytical solution for approximating simple structure in factor analysis. *Psychometrika*, 1953, **18**, 23–38.
[2] Ferguson, G. A. The concept of parsimony in factor analysis. *Psychometrika*, 1954, 19, 281–290.
[3] Holzinger, K. J. and Harman, H. H. *Factor analysis*. Chicago: Univ. Chicago Press, 1941.
[4] Kaiser, H. F. An analytic rotational criterion for factor analysis. *Amer. Psychologist*, 1955, 10, 438. (Abstract)
[5] Kaiser, H. F. Note on Carroll's analytic simple structure. *Psychometrika*, 1956, 21, 89–92.
[6] Kaiser, H. F. The varimax method of factor analysis. Unpublished doctoral dissertation, Univ. California, 1956.
[7] Neuhaus, J. O. and Wrigley, C. The quartimax method: an analytical approach to orthogonal simple structure. *Brit. J. statist. Psychol.*, 1954, 7, 81–91.
[8] Saunders, D. R. An analytic method for rotation to orthogonal simple structure. Princeton: Educational Testing Service Research Bulletin 53–10, 1953.
[9] Thomson, G. H. *The factorial analysis of human ability*. (5th ed.) New York: Houghton Mifflin, 1951.
[10] Thurstone, L. L. *Theory of multiple factors*. Ann Arbor: Edwards Bros., 1932.
[11] Thurstone, L. L. Primary mental abilities. *Psychometric Monogr. No. 1*, 1938.
[12] Thurstone, L. L. *Multiple-factor analysis*. Chicago: Univ. Chicago Press, 1947.

Manuscript received 11/15/57
Revised manuscript received 1/9/58

PSYCHOMETRIKA—VOL. 23, NO. 3
SEPTEMBER, 1958

POWER FUNCTION CHARTS FOR SPECIFICATION OF SAMPLE SIZE IN ANALYSIS OF VARIANCE

LEONARD S. FELDT
STATE UNIVERSITY OF IOWA
AND
MOHARRAM W. MAHMOUD
EGYPTIAN MINISTRY OF EDUCATION

The specification of sample size is an important aspect of the planning of every experiment. When the investigator intends to use the techniques of analysis of variance in the study of treatments effects, he should, in specifying sample size, take into consideration the power of the *F* tests which will be made. The charts presented in this paper make possible a simple and direct estimate of the sample size required for *F* tests of specified power.

A primary consideration in the design of any experiment is the specification of the number of subjects to be selected from the various treatment populations. This number should be such that the important statistical tests will be reasonably sensitive in detecting false null hypotheses. Statistical theory provides the basis for designing such tests; in many psychological and educational experiments sufficient preliminary information is available to permit an application of this theory. The purpose of this paper is to provide power function charts which will simplify the application of the theory and thus facilitate the specification of sample size in experiments employing the techniques of analysis of variance.

The power of the statistical test in any experimental setup—that is, the probability of rejecting the null hypothesis when it is false—depends on the level of significance α at which the test is made, the number of observations or subjects n on which data are available, and the degree of falsity ϕ' of the hypothesis under test. The latter factor is defined as the square root of the ratio of the variance of the treatment population means to the variance for error within the treatment populations. Symbolically,

$$\phi' = \sqrt{\frac{\sum_i^k (\mu_i - \mu)^2/k}{\sigma^2}}.$$

For every F test at a given level of significance in any given design, the power P against any specified alternative to the null hypothesis is uniquely determined by the value of n. Conversely, for every test there exists a value of n which will result in a test of specified power against a specified alternative.

In those experiments in which the power requirements of the F test can be rationally fixed against a specific alternative, it is possible to determine the appropriate sample size. It is for such situations that the present charts are intended.

Nature of the Charts

The charts presented in this paper are for use with tests of the main effects of treatments in experiments involving two to five levels of the treatment variable. The charts are strictly valid only for the completely randomized design; however, they may be applied with relatively little error to tests of treatments effects in randomized block designs and factorial designs employing a within-cells estimate of error variance. A chart presents two families of three curves each. The families pertain to the .05 and .01 levels of significance; the curves within families correspond to power values of .5, .7 and .9.

A separate chart is provided for each value of k, the number of levels of the treatment variable, $f_1 = k - 1$, from 2 through 5. The chart and family appropriate for a given experimental test is entered with the parameter ϕ' along the abscissa. The value of n, the number of observations required per treatment for a test of specified power, is read directly from the ordinate of the chart.

Historical Development

The distribution of the F statistic under hypotheses alternative to the null hypothesis was first considered by Fisher [1] and Wishart [9], who derived expressions for the noncentral F distribution in the form of the correlation ratio. Later Tang [8] derived the same result from the distribution of the noncentral χ^2. Tang also presented extensive tables of the power function. These tables are entered with the parameter ϕ, defined as

$$\phi = \sqrt{\frac{\sum_i^k n(\mu_i - \mu)^2/k}{\sigma^2}}.$$

For fixed values of α, ϕ, f_1, and f_2 the probability of retaining a false null hypothesis may be determined. Unfortunately, the interval of tabulation for ϕ is .50, an interval which is not sufficiently fine for satisfactory interpolation.

Following Tang's procedure, Lehmer [4] tabulated the values of ϕ for $\alpha = .05$ and .01, $P = .7$ and .8 over a wide range of f_1 and f_2. These tables are quite complete within the power range considered; however they cannot be conveniently used in the planning of experiments. From the tables the experimenter can tell only that a projected test will have a power less than .7, between .7 and .8, or greater than .8 against a specified alternative. A greater range of power values would make such tables considerably more useful.

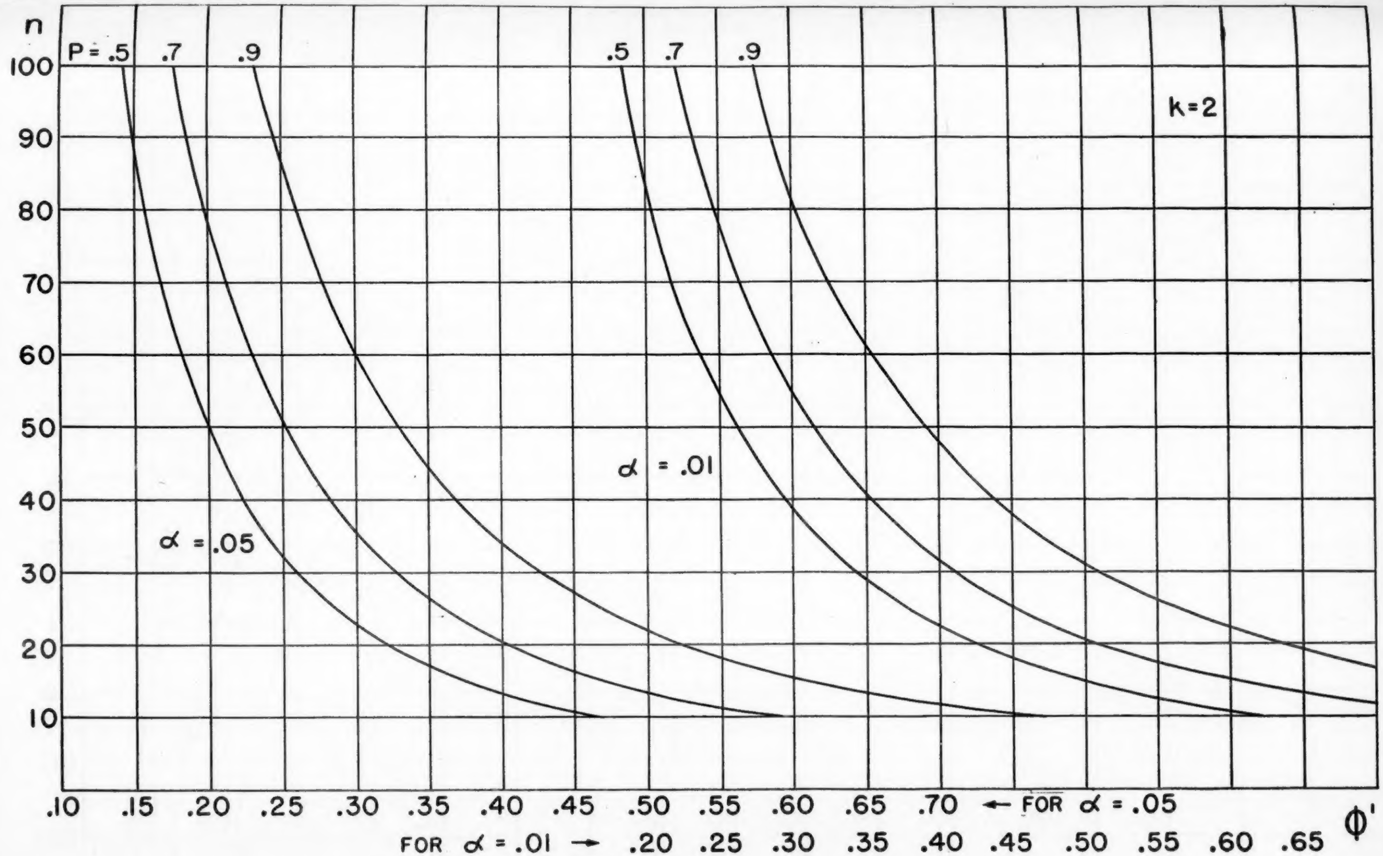

FIGURE 1
Curves of Constant Power for the Test of Main Effects with $k = 2$

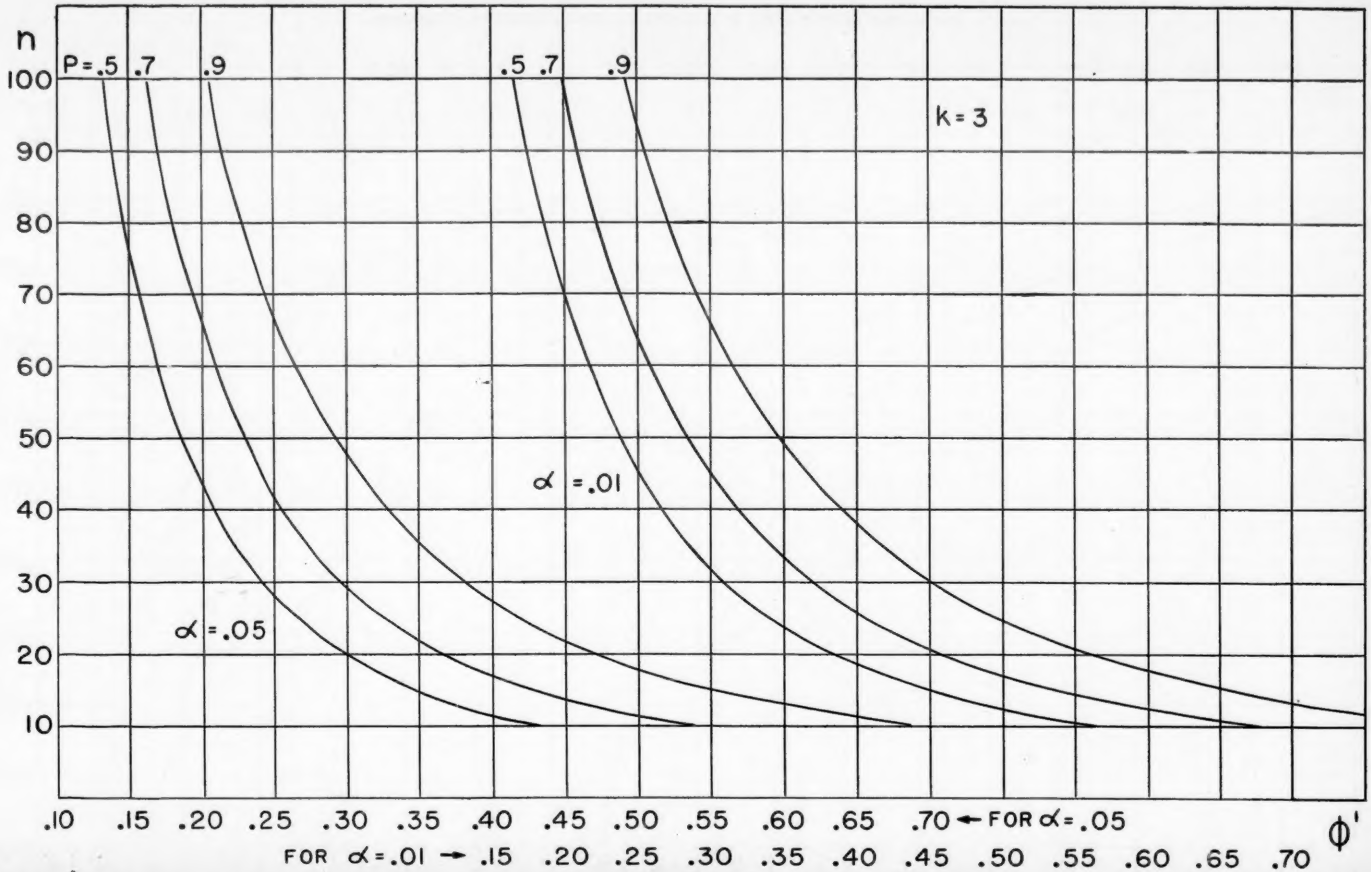

FIGURE 2
Curves of Constant Power for the Test of Main Effects with $k = 3$

FIGURE 2
Curves of Constant Power for the Test of Main Effects with $k = 3$

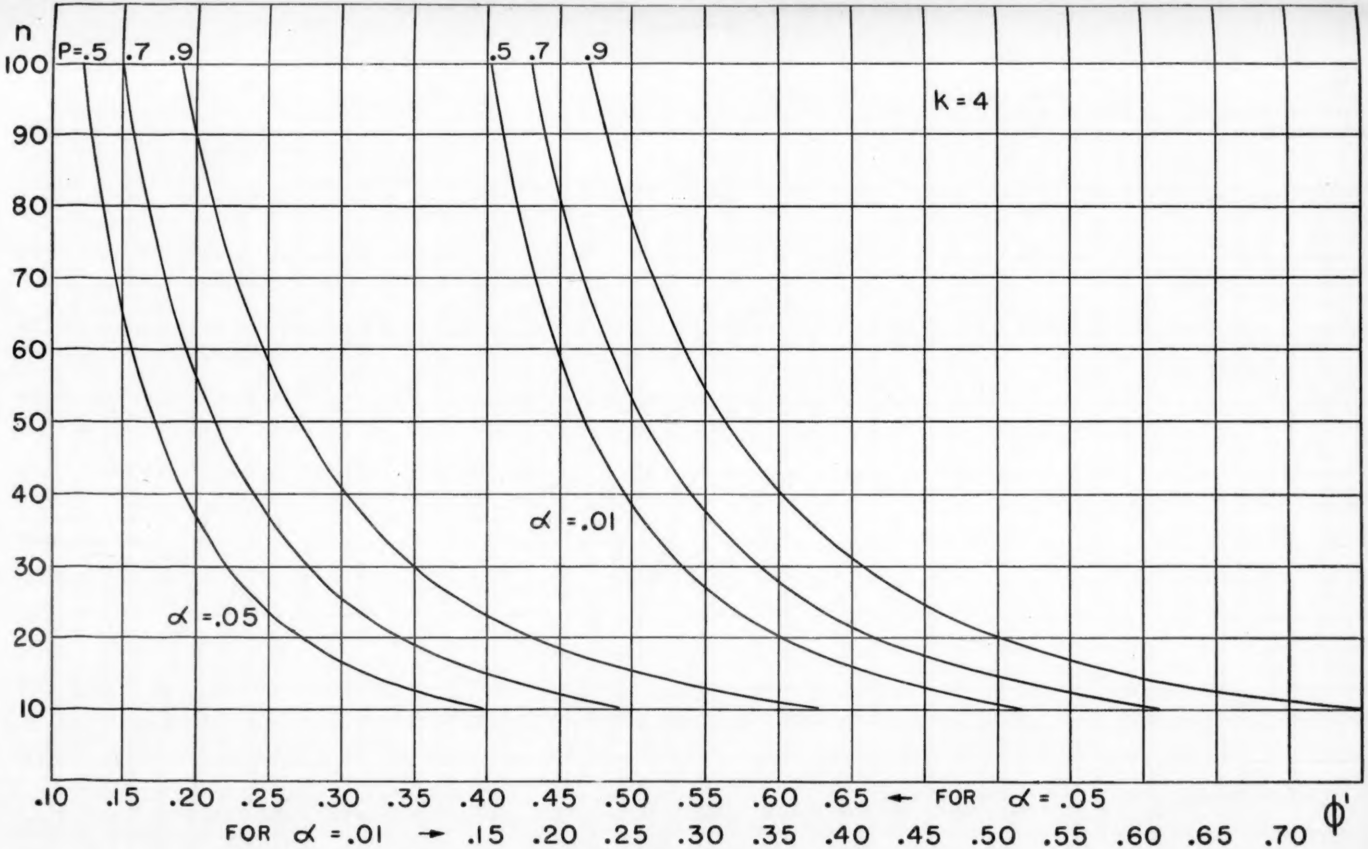

FIGURE 3
Curves of Constant Power for the Test of Main Effects with $k = 4$

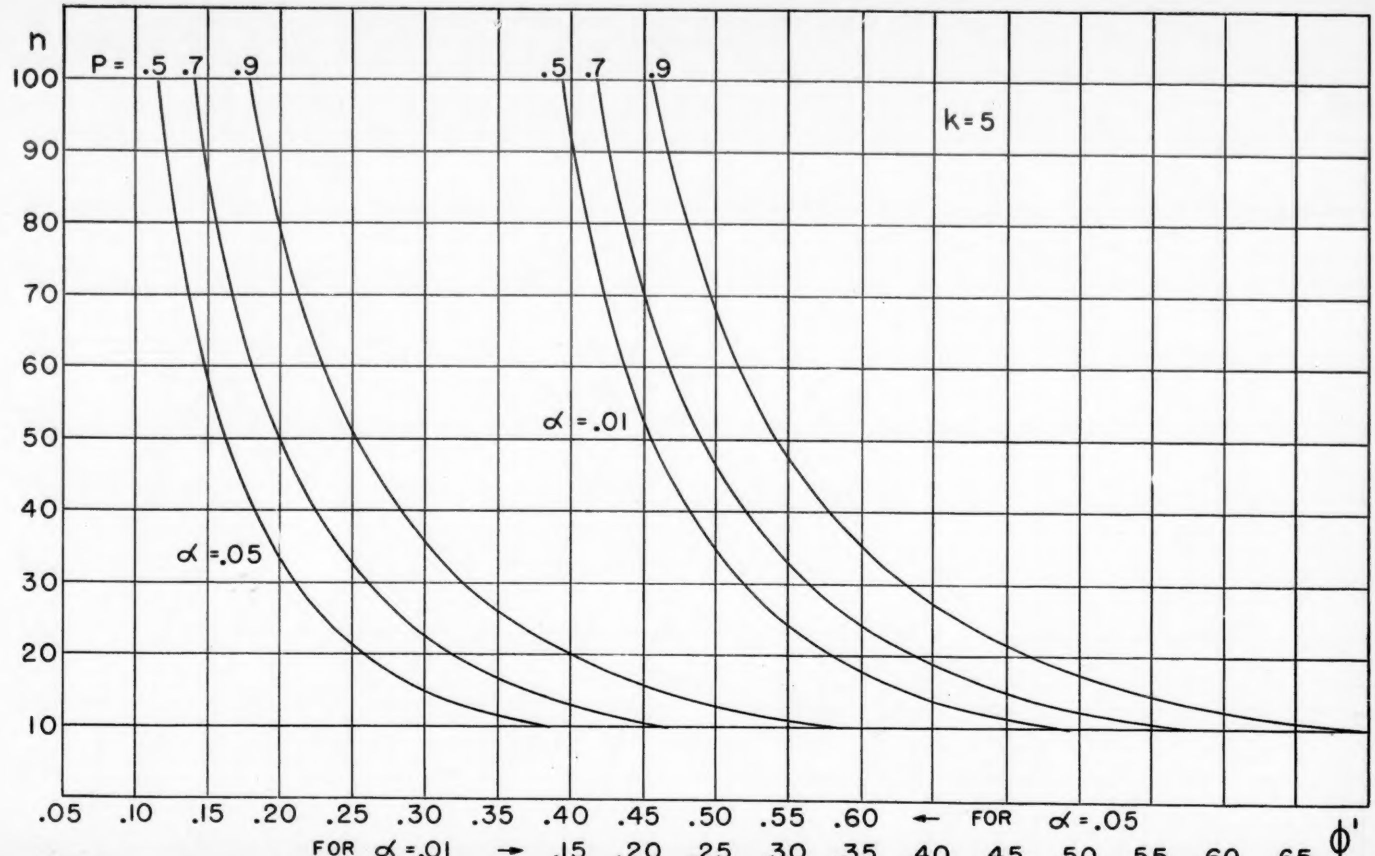

FIGURE 4
Curves of Constant Power for the Test of Main Effects with $k = 5$

Patnaik [6] made an extensive study of the power function of analysis of variance tests. By the method of moments, he derived an approximation to the noncentral F distribution based on the central F distribution. This approximation is computationally feasible but somewhat tedious, especially if a number of power estimates are required. Its primary limitation to psychological experimenters is the labor involved in utilizing Pearson's *Tables of the Incomplete Beta Function* in obtaining power values. This limitation is especially marked in the many instances which demand interpolation within these tables.

Pearson and Hartley [7] presented families of power curves for various combinations of α, f_1, and f_2, which make possible a direct estimate of the power of analysis of variance tests. These curves, like the tables of Tang, are entered with the parameter ϕ. For any given experimental setup, the power of the test may be read directly from the ordinate of the curve. These curves are well suited to the evaluation of the power of any given test. They cannot be easily employed in the inverse manner, however, to indicate the value of n which should be adopted in order to secure a test of specified power. For this purpose, the experimenter must adopt the relatively inefficient approach of making repeated approximations until the value of n has been estimated with sufficient accuracy.

Nicholson [5] and Hodges [3] have derived general formulas for the computation of the power of the analysis of variance test when f_2 is an even number. The formulas involve the evaluation of terms in a certain series, the number of terms being dependent on the number of degrees of freedom for error. The latter feature is a serious practical limitation, for when f_2 is greater than 20, as it often is in psychological experiments, the evaluation becomes too laborious to be of practical utility.

Fox [2] developed charts which overcome some of the objections to earlier works and facilitate the determination of sample size. These charts were constructed from the tables of Tang and Lehmer and are essentially graphs of constant ϕ for varying values of f_1 and f_2. By a method of successive approximations, the value of n may be determined for a fixed value of α and a fixed value of P against a specified alternative hypothesis. These charts are somewhat laborious to apply, however, because of the iterative nature of the approximation for n. Also, the charts do not extend below $f_1 = 3$. For psychological experimenters, who typically deal with fixed treatments effects, this limitation considerably restricts their usefulness.

In theory, the problem of evaluating the power of the test of treatments effects in the simpler analyses of variance has been completely solved. Exact, approximate, and graphical solutions have been derived. However, neither the computational formulas nor the graphical solutions make possible a simple, direct, noniterative approximation of the sample size required for a test of specified power. The charts presented in this paper permit this direct approximation for n, and hence should be of considerable value.

Construction of the Charts

Each chart in this series presents, for $\alpha = .05$ and $.01$, a family of three curves which correspond to $P = .5, .7,$ and $.9$. The numerical calculations for the coordinates of the points on the curve $P = .7$ were carried out from the tables of Lehmer; the calculations for the remaining curves were based on data read from the charts of Pearson and Hartley. The three basic steps in the calculations were as follows.

(1) Determine (from chart or table) pairs of values for ϕ and f_2 for a specified value of P, f_1, and α.

(2) Solve f_2 for n from the relationship $n = 1 + f_2/k$, where k is the number of treatments and n the number of observations per treatment.

(3) Divide ϕ by $\sqrt{n}$ to obtain ϕ'.

The pairs of coordinates for n and ϕ' were then plotted and smooth curves fitted through the points.

Example

An experimenter wishes to compare the level of mastery reached by three groups of college subjects who memorize a list of paired adjectives under three levels of motivation. From a tentative theoretical formulation of the learning task the experimenter predicts the following array of mean differences for the treatment populations at the three levels:

$$M_I - M_{II} = 5.0;$$

$$M_I - M_{III} = 8.0;$$

$$M_{II} - M_{III} = 3.0.$$

Against this alternative to the null hypothesis the experimenter wishes a power of .90 for a test made at the 5 per cent level. Previous experimentation with this list has given rise to an error variance of 100.0, a value which may be taken as a population parameter for this purpose.

From the array of differences the variance of the treatment population means can be computed equal to 10.89. The value of ϕ' is therefore equal to .33. Entering Figure 2 with this value, the required number of subjects is equal to 40.

Note on the Generality of the Charts

The charts presented in this paper are strictly valid only for the test of main effects in the completely randomized design. However, values of n read from these charts underestimate only slightly the values of n which would be required in the randomized block or factorial designs.

The specificity of the charts stems from the unique relationship which holds between n and f_2 in each experimental design. For example, for the completely randomized design (and the one used in the construction of these charts) the relationship is

$$n = 1 + f_2/k.$$

In the randomized block design

$$n = 1 + f_2/(k - 1).$$

For the test of the factor with k levels in the $k \times h$ factorial design (mean square within cells being used as the error term) the relationship is

$$n = h + f_2/k.$$

Because of the differences among these relationships, the numerical relationship of ϕ to ϕ' varies from one design to another, and charts based on that which holds for the completely randomized design will be only approximately correct for the other setups. However, for values of $f_2 \geq 20$, the relationship of ϕ to ϕ' is almost identical for all three designs. We may demonstrate the relatively small error involved in using the present charts for planning randomized block and factorial designs by applying the charts to two examples which tend to maximize the extent of the inaccuracy. This occurs in the randomized block design when $k = 2$; it occurs in the factorial design employing a within-cells error term and proportional frequencies when the number of measures per cell approaches 2. According to Figure 1, a completely randomized design involving two levels of the treatment variable and $\alpha = .05$ will require $n = 11$ ($f_2 = 20$) for $P = .90$ against $\phi' = .725$. The value of n which is actually needed in a randomized block design for $P = .90$ is approximately 12.0. The comparable value for a 2×6 factorial design is 11.9.

This discrepancy, which for even this extreme case is probably of little consequence in the planning of most experiments, is considerably smaller for larger values of f_1 and f_2. Therefore, for practical purposes of approximating the necessary sample size in randomized block and factorial experiments, the tables presented in this paper would seem sufficiently precise.

REFERENCES

[1] Fisher, R. A. The general sampling distribution of the multiple correlation coefficient. *Proc. roy. Soc. London*, 1928, *A*, **121**, 654–673.

[2] Fox, M. Charts for the power of the *F*-test. *Ann. math. Statist.*, 1956, **27**, 484–497.

[3] Hodges, J. L. On the noncentral Beta distribution. *Ann. math. Statist.*, 1955, **26**, 648–653.

[4] Lehmer, E. Inverse tables of probabilities of errors of the second kind. *Ann. math. Statist.*, 1944, **15**, 388–398.

[5] Nicholson, W. L. A computing formula for the power of the analysis of variance test. *Ann. math. Statist.*, 1954, **25**, 607–610.

[6] Patnaik, P. B. The noncentral χ^2 and F-distribution and their applications. *Biometrika*, 1949, **36**, 202–232.
[7] Pearson, E. S. and Hartley, H. O. Charts of the power function for analysis of variance tests, derived from the noncentral F-distribution. *Biometrika*, 1951, **38**, 112–130.
[8] Tang, P. C. The power function of the analysis of variance test with tables and illustrations of their use. *Statist. res. Memoirs*. 1938, **2**, 126–149.
[9] Wishart, J. A note on the distribution of the correlation ratio. *Biometrika*, 1932, **24**, 441–456.

Manuscript received 9/25/57

PSYCHOMETRIKA—VOL. 23, NO. 3
SEPTEMBER, 1958

A COMPARATIVE STUDY OF THREE METHODS OF ROTATION

BENJAMIN FRUCHTER
UNIVERSITY OF TEXAS
AND
EDWIN NOVAK
SYSTEMS DEVELOPMENT CORP.

Three methods of rotation (the graphical, the Thurstone analytical, and the direct-rotational) were applied to the matrix of centroid loadings for 35 variables, to determine which method is the most efficient from theoretical and practical standpoints. The direct-rotational method provided the most information for determining the rank of the configuration and was most economical with respect to time required to reach a rotational solution. The analytical method required the least number of judgmental decisions and was the most objective. The graphical method was the most laborious but had a slight advantage with regard to the number of near-zero loadings in the rotational solution.

Since 1932, the year in which L. L. Thurstone introduced rotational procedures to factor analysis, many different methods of rotation have been developed. A survey of the literature reporting factorial studies has revealed that graphical methods have been employed most frequently to accomplish the rotations even though these methods are relatively slow, laborious, and require numerous skilled judgments. Until the present time, however, they have been the most dependable methods. The majority of the recent proposals for the solution of the rotational problem have been analytical procedures with emphasis on objectivity.

Thurstone [9] and Carroll [1] have developed analytical methods by which solutions may be obtained on orthogonal or oblique axes. Saunders [8] and Neuhaus and Wrigley [7] have developed variants of Carroll's method which yield orthogonal factors. With the exception of Thurstone's method, which requires judgment in selecting the initial weights for a factor, the procedures are sufficiently systematic so that electronic computer programs for them can be written. Differing in approach, a few methods have been advanced that are of particular theoretical interest because they have proposed to reduce the problem to a single rotation. Of these methods, the one offered by Harris [5], based upon geometric models, has shown considerable promise for future development.

This paper reports an investigation which compared three oblique, rotational solutions to a battery of 35 aptitude, achievement, and background variables [2]. The graphical, the Thurstone analytical, and the direct-rotational methods were applied.

The study attempted to assess the validity of the following propositions by comparing the results from the solutions derived by the three rotational methods.

(1) The rank of the factor configuration is determined with greater accuracy by the direct-rotational method than by graphical or analytical methods.

(2) The solution obtained by the direct-rotational method yields a result which is a closer approximation to simple structure than the results obtained by the graphical and analytical methods.

(3) The solution by the direct-rotational method requires less time to reach an approximation to simple structure than do the graphical and analytical solutions.

(4) The direct-rotational solution requires a smaller number of skilled judgments on the part of the investigator than do the graphical or analytical solutions.

Procedure

A comprehensive test battery was administered to 881 male, basic airmen at Lackland Air Force Base, San Antonio, Texas. The mean chronological age of the airmen was 18 years; their mean educational attainment was the eleventh grade.

Product moment correlations were obtained for 35 variables selected from the Army General Classification Tests, the Airmen Classification Test Battery, the Adjutant General Mechanical Aptitude Test, the Differential Aptitude Tests, the Gray-Votaw General Achievement Tests, the Iowa High School Content Examination, the Otis Quick-Scoring Mental Ability Test, the Sims Score Card for Socio-Economic Status, and education in years.

The matrix of intercorrelations was factored by the centroid method. Several empirical criteria, such as Tucker's Phi, Coombs' Criterion, and the range of the centroid loadings indicated the rank of the correlation matrix to be eight. (The intercorrelation and centroid factor matrices have been published in Fruchter [2].)

The Graphical Solution

The graphical method made use of the eight columns of the centroid factor matrix F_c . Each column or axis of F_c was plotted against every other column. Each plot was examined to determine whether a rotation of a pair of columns or axes was indicated. After a pair of axes was rotated, the graphs involving these axes were replotted. The new plots were again inspected to determine the need for additional rotations. The process was continued until the projections on the orthogonal axes approximated simple structure as closely as possible. Since several plots indicated an oblique configuration

TABLE 1

Rotated Factor Loadings of the Solution by the Graphical Method

(All values to three decimal places)

Variables	Factor							
	I V	II N	III Me	IV P	V Vz	VI AI	VII CU	VIII SE
1. AGCT Reading and Vocabulary	376	106	108	072	058	113	012	050
2. AGCT Arithmetic Computations	-009	555	-050	081	116	066	047	-033
3. AGCT Arithmetic Reasoning	132	502	067	-050	232	-017	-059	066
4. AGCT Pattern Analysis	-038	068	006	111	559	-022	067	070
5. AG Mechanical Aptitude 2	-040	066	339	215	201	050	175	026
6. Education in Years	006	197	089	099	-042	182	128	321
7. Arithmetic Reasoning, BI201A	156	443	099	-042	175	031	-033	049
8. Aviation Information, BI101A	130	-003	167	008	093	353	-018	011
9. Background for Current Affairs, BI102A	207	054	013	-015	098	408	-080	005
10. Dial and Table Reading, PB622A-621A	095	296	034	258	235	013	-046	100
11. Electrical Information, BI901A	178	081	413	100	-024	111	048	051
12. General Mechanics, BI902A	112	059	600	-023	-035	050	-005	046
13. Mechanical Principles, BI903A	015	045	374	000	320	-069	149	037
14. Memory for Landmarks, BI510A	118	116	-089	104	378	-009	-017	159
15. Numerical Operations, II, CI702B	149	513	-024	274	-066	-030	-004	-042
16. Reading Comprehension, BI601A	305	127	173	-016	099	139	-003	000
17. Speed of Identification, CP610A	-041	-080	-007	486	231	054	106	-040
18. Tool Functions, BI904A	-100	005	602	113	051	010	-040	001
19. Word Knowledge Form A, BI602A	404	-001	059	-035	076	243	062	024
20. Biographical Inventory, Instructor, BE601B	-083	097	071	062	-050	-019	136	442
21. DAT-Abstract Reasoning, Form A	086	170	-058	009	443	048	-032	094
22. DAT-Clerical, Part II, Form A	097	072	010	570	073	-064	-002	179
23. DAT-Language Usage,Part I, Form A	444	148	044	065	-080	039	191	-062
24. DAT-Language Usage, Part II, Form A	426	100	117	049	041	-081	301	029
25. DAT-Numerical Ability, Form A	008	555	-057	-055	137	106	086	-027
26. DAT-Space Relations, Form A	-061	003	045	067	536	046	073	-025
27. Gray-Votaw, Science	192	071	116	-031	079	262	129	004
28. Gray-Votaw, Social Studies	220	121	001	023	-009	421	003	-056
29. Gray-Votaw, Literature	246	008	-081	012	032	333	115	-055
30. Gray-Votaw, Choice of Words	318	070	056	003	060	059	304	-057
31. Gray-Votaw, Reading	422	060	119	092	081	105	082	-031
32. Gray-Votaw, Arithmetic	097	554	032	056	062	046	015	-058
33. Iowa High School Content, Form L	165	203	000	127	-096	266	216	059
34. Otis Quick-Scoring, Gamma, Form AM	312	184	044	087	174	086	077	025
35. Sims Score Card, Form C	-067	-041	-005	046	078	204	035	500

of factors, and in order to make relative comparisons with the other solutions, an oblique solution was obtained graphically from the orthogonal solution. The oblique solution, in which the factors are identified as verbal, numerical, mechanical, perceptual, visualization, academic information, correct English usage, and socio-economic background, is shown in Table 1, and the cosines of the angular separations of the axes in Table 4.

The apparatus designed by Zimmerman [10], consisting of a drawing board, T-square, and a triangle, was used in the rotational procedures. This eliminated the necessity of converting the projections to numerical values after each rotation and of replotting the numerical values. Eighty-two rotations of pairs of axes were made in obtaining the oblique solution by this method.

The Analytical Solution

Thurstone's analytical procedure is a single-plane method since the coordinate hyperplanes are determined one at a time. The procedure was started by selecting and normalizing test vector 15 (Numerical Operations), which served as a trial vector to find the first reference vector. Test 15, as required in the procedure, has some relatively high and low correlations in the correlation matrix. The trial vector was then adjusted to a position Λ_I which is the normal that defines the first hyperplane.

The test vector 31 (Gray-Votaw Reading), which has a low projection on the first reference vector Λ_I, was selected to become the second trial vector. (A projection was considered low when its absolute value was low relative to the rest of the test projections on the same reference vector.) It was employed to determine reference vector Λ_{II}. The test vector 18 (Tool Functions) with low projections on Λ_I and Λ_{II} was selected as the third trial vector to find reference vector Λ_{III}.

This process was continued until the number of reference vectors so determined was equal to the number of factors in the centroid factor matrix. The eight factors identified by the analytical solution were the same as those identified in the graphical solution, although two of the factors, correct English usage and academic information, did not appear to be independent of the verbal factor according to the requirements for simple structure. The rotated factor matrix obtained by the analytical solution is given in Table 2 and the cosines of the angles between the axes in Table 5.

The Direct-Rotational Solution

The method outlined by Harris [4, 5] was used to obtain the third solution. In this method the assumption is made that a given factor configuration is consistent with a specific geometric model. The postulated figure is a tetrahedron for 3-factor problems while the hypertetrahedron is the analogous model for higher-dimensional problems. The objective is to

TABLE 2

Rotated Factor Loadings of the Solution by Thurstone's Analytical Method

(All entries to three decimal places.)

Variables	Factor I N	II V	III Me	IV Vz	V SE	VI P	VII AI	VIII CU
1. AGCT Reading and Vocabulary	046	411	034	082	035	-019	347	290
2. AGCT Arithmetic Computations	514	-090	-020	272	-011	045	014	-061
3. AGCT Arithmetic Reasoning	308	000	043	417	008	-124	016	002
4. AGCT Pattern Analysis	-050	-070	054	457	012	265	-055	-017
5. AG Mechanical Aptitude 2	045	028	394	101	016	292	159	078
6. Education in Years	178	088	064	-001	373	022	175	062
7. Arithmetic Reasoning, BI201A	274	070	074	337	011	-113	093	051
8. Aviation Information, BI 101A	-024	304	184	126	096	044	299	123
9. Background for Current Affairs, BI102A	037	367	011	092	107	-008	128	305
10. Dial and Table Reading, BP622A-621A	284	017	048	258	048	190	-009	-014
11. Electrical Information, BI901A	038	255	394	-025	042	034	301	190
12. General Mechanics, BI902A	-054	168	589	-008	004	-084	234	133
13. Mechanical Principles, BI903A	-129	023	397	293	-033	103	131	118
14. Memory for Landmarks, BI510A	025	059	-109	356	098	117	-025	040
15. Numerical Operations, II, CI702B	561	023	-038	084	-066	105	047	020
16. Reading Comprehension, BI601A	023	350	126	137	-005	-060	319	237
17. Speed of Identification, CP610A	105	012	069	017	-038	567	021	006
18. Tool Functions, BI904A	-026	-062	664	003	-048	116	021	-056
19. Word Knowledge Form A, BI602A	-082	533	-014	064	066	-056	477	368
20. Biographical Inventory, Instructor, BE601B	055	-075	024	-016	428	-030	004	000
21. DAT-Abstract Reasoning, Form A	027	042	-054	443	052	071	-001	012
22. DAT-Clerical, Part II, Form A	262	039	002	-027	107	437	-033	015
23. DAT-Language Usage, Part I, Form A	103	478	-052	-036	-060	-022	490	437
24. DAT-Language Usage, Part II, Form A	-021	439	009	047	-017	005	494	493
25. DAT-Numerical Ability, Form A	421	-043	-038	314	016	-056	084	-014
26. DAT-Space Realtions, Form A	-102	-041	116	408	-049	268	-006	-011
27. Gray-Votaw, Science	000	338	103	066	079	010	392	242
28. Gray-Votaw, Social Studies	136	397	002	002	-041	019	389	179
29. Gray-Votaw, Literature	015	423	-100	-009	061	055	424	268
30. Gray-Votaw, Choice of Words	-023	403	-002	038	-033	042	491	423
31. Gray-Votaw, Reading	005	479	049	066	-043	041	434	373
32. Gray-Votaw, Arithmetic	470	007	036	243	-057	-024	084	011
33. Iowa High School Content, Form L	235	303	053	-074	167	095	401	234
34. Otis Quick-Scoring, Gamma, Form AM	099	326	005	193	027	053	307	260
35. Sims Score Card, Form C	-063	042	-037	039	541	004	038	-025

locate the primary axes at the intersections of the hyperplanes. The loadings of points at the intersections of the hyperplanes provide direction numbers which constitute a basis for deriving the transformation matrix. Once these points are established, a single rotation is required to obtain the rotated loadings of the variables on either primary or simple axes.

The test vectors of the factor matrix F_c were extended to the mean of the first centroid loadings rather than to unity, as is usually done, in an effort to reduce distortion introduced by extending vectors with low values on the first centroid. The extended loadings were plotted on combinations of pairs of axes. The plots were inspected to establish the locations of the eight primary axes. The same factors were identified in this solution as in the graphical and analytical solutions. However, the high intercorrelations of the verbal, academic interest, and correct English usage factors plus several extremely high values in the inverse of the intercorrelation matrix of the eight primary factors indicated that the axes were not sufficiently independent to represent separate factors. The possibility of a seven-factor solution in which the verbal and correct English usage factors of the previous solution were combined was investigated next. The two factors in the verbal domain, verbal and academic interest, were still too highly correlated in this solution to be considered separate dimensions.

A solution based on the assumption that only six independent factors were present in the test battery was then tried and accepted. This solution, which is presented in Table 3, satisfies the criteria for determining the rank of a factor configuration. Briefly, the criteria stipulate that: (1) the extended vector configuration should be consistent with the assumed geometric model; (2) the intercorrelations of the factors should be small enough to give evidence of their independence; and (3) the inverse of the intercorrelation matrix of the factors should be computed satisfactorily, i.e., most of the entries should be small, preferably less than unity. The cosines of the angles of the six factors which were identified as verbal, numerical, mechanical, perceptual, visualization, and socio-economic background are given in Table 6.

Results

The rank of the factor configuration by the graphical method was judged to be eight. The sole criterion applied during rotation specified that, if one or more of the factors achieved only small loadings (i.e., between $\pm$ .20), it would be residualized and the rank would have been less than eight. The correct English usage factor was not identified with confidence, although it did not residualize. The application of simple structure criteria and an inspection of a set of plots of the rotated factors served as a final check on the adequacy of the rotational solution.

TABLE 3

Rotated Factor Loadings of the Solution by the Direct-Rotational Method

(All entries to three decimal places.)

Variables	Factor I V	II N	III Me	IV P	V Vz	VI SE
1. AGCT Reading and Vocabulary	382	118	110	029	016	000
2. AGCT Arithmetic Computations	-054	569	-059	-123	203	144
3. AGCT Arithmetic Reasoning	006	493	016	-228	319	124
4. AGCT Pattern Analysis	-038	070	-032	019	448	059
5. AG Mechanical Aptitude 2	061	067	346	135	128	048
6. Education in Years	053	232	115	-006	-050	225
7. Arithmetic Reasoning, BI201A	074	431	065	-196	256	105
8. Aviation Information, BI101A	333	-039	184	-004	085	102
9. Background for Current Affairs, BI102A	414	028	027	-039	099	122
10. Dial and Table Reading, BP622A-621A	-010	349	-004	096	157	146
11. Electrical Information, BI901A	203	070	416	055	-035	038
12. General Mechanics, BI902A	104	008	582	-045	011	011
13. Mechanical Principles, BI903A	010	003	361	-047	294	-032
14. Memory for Landmarks, BI510A	058	148	-130	011	280	116
15. Numerical Operations, II, CI702B	-008	582	-043	077	-049	068
16. Reading Comprehension, BI601A	344	106	171	-056	106	-012
17. Speed of Identification, CP610A	001	-002	-004	419	008	007
18. Tool Functions, BI904A	-120	-032	564	077	046	022
19. Word Knowledge Form A, BI602A	542	-018	092	-026	047	-016
20. Biographical Inventory, Instructor, BE601B	-095	148	081	-013	-080	362
21. DAT-Abstract Reasoning, Form A	075	167	-101	-090	397	102
22. DAT-Clerical, Part II, Form A	-059	210	-025	443	-160	146
23. DAT-Language Usage, Part I, Form A	455	168	095	035	-094	-135
24. DAT-Language Usage, Part II, Form A	398	117	168	021	-006	-140
25. DAT-Numerical Ability, Form A	016	539	-052	-239	268	143
26. DAT-Space Relations, Form A	015	-022	022	010	446	000
27. Gray-Votaw, Science	368	040	160	-054	090	046
28. Gray-Votaw, Social Studies	448	098	043	-013	012	093
29. Gray-Votaw, Literature	475	-004	-019	015	013	013
30. Gray-Votaw, Choice of Words	414	061	127	-007	041	-129
31. Gray-Votaw, Reading	459	068	138	067	020	-096
32. Gray-Votaw, Arithmetic	025	561	018	-134	162	085
33. Iowa High School Content, Form L	313	229	063	048	-094	143
34. Otis Quick-Scoring, Gamma, Form AM	321	197	049	007	131	002
35. Sims Score Card, Form C	003	-001	-001	002	000	465

TABLE 4

Cosines of Angular Separations Between Simple Axes
Graphical Solution
(All entries to three decimal places)

	I V	II N	III Me	IV P	V Vz	VI AI	VII CU	VIII SE
I	1000							
II	-246	1000						
III	095	-040	999					
IV	092	-234	106	1000				
V	-094	-138	-433	-143	1000			
VI	-425	-066	-299	-258	-104	1000		
VII	006	-085	009	051	-120	-196	1000	
VIII	-160	-043	063	009	098	-101	-036	1000

TABLE 5

Cosines of Angular Separations Between Simple Axes
Thurstone Analytical Method
(All entries to three decimal places)

	I N	II V	III Me	IV Vz	V SE	VI P	VII AI	VIII CU
I	1000							
II	-296	1000						
III	111	-025	1000					
IV	046	-157	-236	1000				
V	080	-059	051	152	1000			
VI	166	024	104	-148	003	1000		
VII	-153	819	194	-440	011	096	1000	
VIII	-353	831	022	-195	-144	125	872	1000

TABLE 6

Cosines of Angular Separations Between Simple Axes

Direct Rotational Method

(All entries to three decimal places)

	I V	II N	III Me	IV P	V Vz	VI SE
I	1000					
II	-494	1000				
III	-033	-112	1000			
IV	086	-328	136	1000		
V	-172	097	-467	-628	1000	
VI	-520	278	-176	-263	113	1000

The analytical solution identified the same eight factors. However, the application of the simple structure criteria placed the correct English usage and academic information factors in a doubtful category [3]. The similarity of the verbal, correct English usage, and academic information factors was indicated by a questionable fulfilment of the requirement that for every pair of factors there should be several variables with projections on one factor vector but not on the other. There was also a question as to whether the number of tests was small enough to conform to the requirement that for every pair of factors only a small number of variables should have appreciable loadings on any pair of factors.

Three separate solutions, based on the assumptions that the rank of the factor configuration was eight, seven, and six, respectively, were obtained by the direct-rotational method. The hypothesis regarding six factors was accepted because the solution produced a figure consistent with the postulated geometric model; the intercorrelation matrix of the six factors gave evidence of their independence; and the inverse of the intercorrelation matrix of the six factors was computed satisfactorily.

The disagreement in the rank of the factor configuration may be attributed to the restriction imposed by the direct-rotational method in regard to the location of reference axes. The criteria were strict and tended to determine conservatively a rank characterized by highly independent factors. On the other hand, the graphical method permitted more latitude in locating the reference axes of the verbal domain, so long as the simple structure criteria were not violated. The placement of reference axes was somewhat more objective by the analytical method. One reference axis or factor was located at a time, and the rank was determined, in the final analysis, by the

application of the simple structure criteria. The rank of the factor configuration seemed to be most precisely determined by the direct-rotational method since criteria for the locations and angular separations of the primary axes were more definitely specified by the additional requirements.

The comparison of the six factors common to the three solutions indicated no appreciable differences in approximating simple structure. The largest number of vanishingly small loadings ($a_{ji} = \pm.20$) on each of the six factors was used as a discriminant. The graphical solution had a slight edge over the other two solutions in this respect.

The solution obtained by the direct-rotational method was most economical with respect to amount of time. It required 88 man-hours to reach a solution which was approximately 30 and 50 hours less than the analytical and graphical solutions respectively.

The analytical method was least demanding of judgmental decisions on the part of the rotater. Judgment was required in the selection of a variable to serve as the initial trial vector for each factor. The remainder of the method consists of routine statistical calculations. The major judgmental portion of the direct-rotational method consisted of locating the intersections of the hyperplanes. This portion was concentrated near the beginning of the procedure, with the rest of the method consisting of routine statistical calculations. The graphical method required a judgment for each rotation of a pair of axes and was most demanding in this respect.

Conclusions

The direct-rotational method, in batteries for which it is suitable (i.e., batteries with a central factor on which all the variables have at least moderate positive loadings), provides a better basis for determining the rank of the factor configuration and is more economical with respect to time required to reach a rotational solution than the other two methods. The analytical method is most useful where it is desired to hold judgmental decisions to a minimum and to turn the calculations over to a statistical clerk or computer. The graphical method is the most laborious but had a slight advantage in the present study with regard to the number of near-zero loadings in the rotational solution.

REFERENCES

[1] Carroll, J. B. An analytical solution for approximating simple structure in factor analysis. *Psychometrika*, 1953, **18**, 23–38.

[2] Fruchter, B. Orthogonal and oblique solutions to a battery of aptitude, achievement, and background variables. *Educ. psychol. Measmt*, 1952, **12**, 20–38.

[3] Fruchter, B. *Introduction to factor analysis*. New York: Van Nostrand, 1954.

[4] Harris, C. W. Direct rotation to primary structure. *J. educ. Psychol.*, 1948, **39**, 449–468.

[5] Harris, C. W. Projections of three types of factor patterns. *J. exp. Educ.*, 1949, **17**, 335–345.

[6] Harris, C. W. and Knoell, D. M. The oblique solution in factor analysis. *J. educ. Psychol.*, 1948, **39**, 385–403.

[7] Neuhaus, J. O. and Wrigley, C. The quartimax method: an analytical approach to orthogonal simple structure. *Brit. J. statist. Psychol.*, 1954, **7**, 81–91.

[8] Saunders, D. R. An analytical method for rotation to orthogonal simple structures. Educational Testing Service Research Bulletin, 53-10, August 1953.

[9] Thurstone, L. L. An analytical method for simple structure. *Psychometrika*, 1954, **19**, 173–182.

[10] Zimmerman, W. S. A simple graphical method for orthogonal rotation of axes. *Psychometrika*, 1946, **11**, 51–55.

Manuscript received 9/23/57
Revised manuscript received 1/6/58

PSYCHOMETRIKA—VOL. 23, NO. 3
SEPTEMBER, 1958

ANALYSIS OF VARIANCE FOR CORRELATED OBSERVATIONS

RAYMOND O. COLLIER, JR.
UNIVERSITY OF MINNESOTA

Two different linear models are presented for the four-dimensional classification system in which correlations exist between certain pairs of observations. Except for the assumption of correlated observations, classical assumptions associated with classification systems are made. The models considered are modifications of those which underlie the split-plot design and the split-split-plot design. In the first model the correlations between observations of the levels of one dimension are all set equal to ρ. In the second model the observations of the levels of one dimension are assumed correlated to degree ρ_1, whereas the observations of a second dimension are correlated to degree ρ_2. Analyses for the two models and tests of hypotheses for various parameters are indicated.

In the social sciences it is generally agreed that the analysis of data involving several observations on a subject is of basic interest and importance. Typically, several different variables are observed for every experimental subject. A special case arises when several repeated or multiple observations are made on a single variable for every individual subject, each successive observation usually being associated with a different level or combination of dimensions of the design. In both kinds of settings it is usual to assume that the observations associated with a single subject are correlated.

Concern over such experiments involving repeated measurements was manifested quite early in statistics and the social sciences. Recent consideration of this problem has been made by several writers. Kogan [9] outlined an analysis for a design in which several groups of subjects were administered different experimental treatments with each individual subject being measured on several trials. Edwards [4] devoted a chapter to the analysis of designs which involved repeated measurements on the same subjects. Later, in surveying the status of experimental design in psychology, Kogan [10] pointed out the relevance of the split-plot design in experiments for which repeated observations were taken on each subject. Perhaps the most comprehensive treatment of the design and analysis of experiments involving repeated measurements for two- and three-dimensional setups has been given by Lindquist [12] in his chapter on mixed designs. An early example of the application of the split-plot principle in an experiment on drawing involving repeated or correlated observations was given by Hoyt, Stunkard, *et al.* [7]. Moonan [13, 14], in considering test reliability and experimental design, presented a model which involved observations assumed to be correlated.

Perhaps the first presentations of the analyses of designs in which correlations between observations could be assumed were made by Yates

[18, 19, 20]. In these papers Yates presented the analysis of the split-plot design as it applied to agricultural experimentation. Later, Nandi [15] supplied a transformation useful in the derivations of possible tests of hypotheses in split-plot designs. The work of Halperin [5] parallels that of Nandi, the pattern of correlation coefficients assumed by both writers being similar. Notable among the many books treating the split-plot design and variations thereof are those by Cochran and Cox [2] and Kempthorne [8]. In addition to the above, a more comprehensive list of references pertaining to the split-plot design and the analysis of designs with correlated observations has been made by the author [3].

With few exceptions work relative to the split-plot design has been done by workers in fields other than the social sciences. The result has been that the ramifications of this design and certain variations of it are not greatly appreciated by workers in the social sciences. Moreover, the simple split-plot design as usually presented, may be generalized to higher dimensional designs only if a good deal of analysis of variance insight and intuition is employed. Finally, it has been practically impossible to find the analyses of such generalized designs presented in the literature, let alone the underlying assumptions listed.

It is the purpose of this paper to study a class of designs in which data of the sort mentioned above are involved. The analyses included will fall into the general setting of the analysis of variance as applied to experimental design.

A Four-Dimensional System with Constant Correlations Between Levels of One Dimension

It should be sufficient for many investigations to consider layouts in which there is a maximum of four dimensions and equal frequencies in the subclasses. The different dimensions may refer to experimental treatments or simply to ways of classification. Suppose that limitations on the possible number of subjects available, constraints imposed by scheduling, or the desires or interests of the experimenter, require that multiple measurements be taken on each individual subject of a sample. Suppose further that each of these observations refers to an observation made under a level of a particular dimension of the design. If the dimension is experimental, each subject is customarily presented the treatments in random order so as to balance sequence effects. Under such circumstances it will be possible to make certain assumptions and arrive at various tests of hypotheses.

Let X_{qrstu} stand for the variable observed on the qth individual for the rth level of dimension r, the sth level of dimension s, the tth level of dimension t, and the uth level of dimension u. Let the dimension to which subscript u refers be that which involves correlated measures. Each subject is observed under all of the levels of u, although under only one of the combi-

nations of r, s, and t. The subscripts on the random variables, X_{qrstu}, will be simple to follow if $q = 1, 2, \cdots, Q$; $r = 1, 2, \cdots, R$; $s = 1, 2, \cdots, S$; $t = 1, 2, \cdots, T$; and $u = 1, 2, \cdots, U$. Thus, the $QRST$ subjects, i.e., Q different subjects in each of RST subclasses or dimension level combinations, are measured on each of U levels of one dimension.

As previously stated assume that correlations exist between pairs of observations taken on the same subject. One model which may be utilized assumes that the random variables X_{qrstu} are $QRSTU$-variate normally distributed with variances and covariances as follows.

$$(1) \qquad V[X_{qrstu}] = \sigma^2;$$

$$\text{cov}\,[X_{qrstu}, X_{q'r's't'u'}]$$
$$= \rho\sigma^2, \quad \text{for} \quad q = q', \quad r = r', \quad s = s', \quad t = t', \quad u \neq u',$$
$$= 0 \quad \text{otherwise.}$$

Let the expected value of a single observation, given as a linear function of fixed constants, be identical to that usually expressed in Model I of the analysis of variance.

$$(2) \qquad E(X_{qrstu}) = \mu + \alpha(1, r) + \alpha(2, s) + \alpha(3, t) + \alpha(4, u) + \beta(1, rs)$$
$$+ \beta(2, rt) + \beta(3, st) + \beta(4, ru) + (\beta5, su) + \beta(6, tu)$$
$$+ \gamma(1, rst) + \gamma(2, rsu) + \gamma(3, rtu) + \gamma(4, stu) + \delta(rstu)$$

where the fixed constants are defined as follows:

- μ refers to the general effect;
- α refers to the main effects of the different dimensions indicated; the numerals being employed to distinguish the main effects of different dimensions when specific values of r, s, t, or u are considered;
- β refers to the first-order interactions between the two dimensions indicated;
- γ refers to the second-order interactions between the three dimensions indicated; and
- δ refers to the third-order interaction constants between all four dimensions.

Since the fixed model is assumed, the conditions concerning the distribution of the X_{qrstu} are equivalent to

$$X_{qrstu} = E(X_{qrstu}) + \epsilon_{qrstu},$$

where the ϵ_{qrstu} refer to random effects. The random components, ϵ_{qrstu}, are assumed to be $QRSTU$-variate normally distributed with

$$(3) \qquad E(\epsilon_{qrstu}) = 0,$$
$$V(\epsilon_{qrstu}) = \sigma^2,$$

and

$$\text{cov}\,(\epsilon_{qrstu}\,,\,\epsilon_{q'r's't'u'}) = \rho\sigma^2, \text{ for } q = q',\, r = r',\, s = s',\, t = t',\, u \neq u'$$
$$= 0 \text{ otherwise.}$$

In addition, the following parametric restrictions are imposed:

$$\sum \alpha = \sum \beta = \sum \gamma = \sum \delta = 0.$$

Thus the sum of parameters over a parscript involved is zero.

In effect, the assumptions made on the variances and correlations state that the variance of the random observation X_{qrstu} is σ^2, the correlations between any two observations made on the same subject are constantly ρ, but the correlations between two observations not made on the same subject are zero, i.e., these observations are stochastically independent.

It is possible, by means of an orthogonal transformation, to transform the original variables X_{qrstu} to two sets of variates which are independent within a set and between sets and which have constant variances within a set, but possess variances which differ from set to set. This transform, due to Nandi [15], follows.

$$Y_{qrstw} = \sqrt{\frac{w}{w+1}} \left\{ X_{qrst(w+1)} - \frac{\sum_{u=1}^{w} X_{qrstu}}{w} \right\},$$

(4) $$(w = 1, 2, \cdots, U - 1).$$

$$Y_{qrstU} = \sqrt{U} \left\{ \frac{\sum_{u=1}^{U} X_{qrstu}}{U} \right\}.$$

From this transformation, in set (1) there are $QRST(U - 1)$ variates Y_{qrstu} $(u = 1, 2, \cdots, U - 1)$, which are normally and independently distributed with constant variance $\sigma_1^2 = \sigma^2 (1 - \rho)$. If the Y_{qrstU} are labeled Z_{qrstU}, it follows that in set (2) there are $QRST$ variates, Z_{qrstU}, normally independently distributed with constant variance $\sigma_2^2 = \sigma^2 [1 + (U - 1)\rho]$, which are also independent of variates in set (1). Since the transformation of (4) is orthogonal it also follows immediately that the sum of squares associated with the variates of set (1), the Y_{qrstu} $(u = 1, 2, \cdots, U - 1)$, is

(5) $$\sum_q \sum_r \sum_s \sum_t \sum_{u=1}^{U-1} [Y_{qrstu} - E(Y_{qrstu})]^2$$

$$= \sum_q \sum_r \sum_s \sum_t \sum_u [X_{qrstu} - \frac{X_{qrst.}}{U} - \alpha(4, u) - \beta(4, ru) - \beta(5, su)$$
$$- \beta(6, tu) - \gamma(2, rsu) - \gamma(3, rtu) - \gamma(4, stu) - \delta(rstu)]^2,$$

and that the sum of squares associated with the variates of set (2), Z_{qrstU}, is

$$\sum_q \sum_r \sum_s \sum_t [Z_{qrstU} - E(Z_{qrstU})]^2 \tag{6}$$

$$= U \sum_q \sum_r \sum_s \sum_t \left[\frac{X_{qrst.}}{U} - \mu - \alpha(1, r) - \alpha(2, s) - \alpha(3, t) - \beta(1, rs) - \beta(2, rt) - \beta(3, st) - \gamma(1, rst)\right]^2,$$

where, in the above expressions, $X_{qrst.} = \sum_{u=1}^{U} X_{qrstu}$ and any sum expressed without limits of summation is to be interpreted as a sum over the entire range of the subscript.

Noting that the parameters in (5) are involved there only, and a similar statement is true of (6), normal regression theory, e.g., as presented by Wilks [17] or one of other equivalent techniques may be utilized to derive tests of hypotheses on parameters involved in the two variate sets. The quantities necessary to test these hypotheses are presented in the analysis of variance of Table 1 and the notation employed in the specification of the test statistics follows this table.

Tests of hypotheses are carried out as is usual in the analysis of the split-plot design. That is, there is an error term for each of the two sets which is used to test hypotheses in a particular set. For example, the test of the hypothesis, $H_0 : \delta(rstu) = 0$, in set (1) is made by forming a ratio of the mean square related to the hypothesis and the mean square due to error (1),

$$F(\delta) = \frac{\text{S.S.}[\delta(rstu)]/(R - 1)(S - 1)(T - 1)(U - 1)}{\text{S.S.(Error 1)}/(Q - 1)(RST)(U - 1)} \tag{7}$$

where S.S. $[\delta(rstu)]$ and S.S. (Error 1) refer to sums of squares appearing in Table 1. $F(\delta)$, under H_0, is distributed as Snedecor's F distribution with $(R - 1)$ $(S - 1)$ $(T - 1)$ $(U - 1)$ and $(Q - 1)$ (RST) $(U - 1)$ degrees of freedom. Therefore, to test H_0 above, with a significance level of α, require, as is usual, that $F[\delta]$ exceed the upper α-point of this particular F distribution.

Tests of other hypotheses in set (1) are made by forming analogous F ratios using the denominator as in (7) above (see [8] pp. 91, 375). Concerning the pooling problem the reader should refer to a discussion of some of the consequences of pooling nonsignificant interaction effects with error recently given by Binder [1]. It should also be noted that a testing procedure recently proposed by Hartley [6] seems promising here and elsewhere in this paper. Tests of hypotheses on parameters involved in set (2) follow perfectly analogous lines. Thus, as an example, the test of the hypothesis, $H_0 : \gamma\,(1, rst) = 0$, is made by means of the F ratio,

$$F[\gamma(1)] = \frac{\text{S.S.}[\gamma(1, rst)]/(R - 1)(S - 1)(T - 1)}{\text{S.S.(Error 2)}/(Q - 1)(RST)}. \tag{8}$$

It is worth noting that the final result has been that of dividing the total sum of squares into two portions, namely the sum of squares associated with Y_{qrstu}, $u = 1, 2, \cdots, U-1$, given by

$$\sum_q \sum_r \sum_s \sum_t \sum_u X^2_{qrstu} - \sum_q \sum_r \sum_s \sum_t X^2_{qrst.}/U,$$

(which is seen to be the sums of squares between observations on the same individual summed over all individuals) and the uncorrected sum of squares between individuals,

$$\sum_q \sum_r \sum_s \sum_t X^2_{qrst.}/U.$$

TABLE 1

Analysis of Variance for a Four Dimensional Layout with Constant Correlations Between the Observations of Levels of One Dimension

Source of Variation	Degrees of Freedom	Sum of Squares
Total Set(1)	QRST(U-1)	a'-b'
(Within individuals)		
$\alpha(4,u)$	U-1	o-p
$\beta(4,ru)$	(R-1)(U-1)	i-ℓ-o+p
$\beta(5,su)$	(S-1)(U-1)	j-m-o+p
$\beta(6,tu)$	(T-1)(U-1)	k-n-o+p
$\gamma(2,rsu)$	(R-1)(S-1)(U-1)	c-f-i-j+ℓ+m+o-p
$\gamma(3,rtu)$	(R-1)(T-1)(U-1)	d-g-i-k+ℓ+n+o-p
$\gamma(4,stu)$	(S-1)(T-1)(U-1)	e-h-j-k+m+n+o-p
$\delta(rstu)$	(R-1)(S-1)(T-1)(U-1)	a-b-c-d-e+f+g+h+i+j+k-ℓ-m-n-o+p
Error(1)	(Q-1)(RST)(U-1)	a'-b'-a+b
Total Set(2)	QRST	b'
(Between individuals)		
μ	1	p
$\alpha(1,r)$	R-1	ℓ-p
$\alpha(2,s)$	S-1	m-p
$\alpha(3,t)$	T-1	n-p
$\beta(1,rs)$	(R-1)(S-1)	f-ℓ-m+p
$\beta(2,rt)$	(R-1)(T-1)	g-ℓ-n+p
$\beta(3,st)$	(S-1)(T-1)	h-m-n+p
$\gamma(1,rst)$	(R-1)(S-1)(T-1)	b-f-g-h+ℓ+m+n-p
Error(2)	(Q-1)(RST)	b'-b

Similarly, the error sum of squares in set (1) may be observed to be the sum of the interactions of individual and dimension u, whereas the error sum of squares in set (2) is simply the sum of squares between individuals in a subclass summed over subclasses. The sums of squares associated with the various effects are computed in the usual fashion.

The parameters listed in the "Source of Variation" column of Table 1 should be interpreted as referring to the source of variation associated with the test of hypothesis on the parameters of the effect being considered. Thus $\alpha(4, u)$ refers to the source of variation connected with $H_0 : \alpha(4, u) = 0$.

The mean squares have not been given in Table 1 since they are computed in the usual fashion. The expected mean squares for error (1) and error (2) may be shown to be

$$\sigma_1^2 = \sigma^2(1 - \rho) \quad \text{and} \quad \sigma_2^2 = \sigma^2[1 + (U - 1)\rho] ,$$

respectively.

In Tables 1 and 2 the notation employed is given by

$$\begin{aligned}
(9) \quad a &= \sum_r \sum_s \sum_t \sum_u X^2_{\cdot rstu}/Q, & b &= \sum_r \sum_s \sum_t X^2_{\cdot rst\cdot}/QU, \\
c &= \sum_r \sum_s \sum_u X^2_{\cdot rs\cdot u}/QT, & d &= \sum_r \sum_t \sum_u X^2_{\cdot r\cdot tu}/QS, \\
e &= \sum_s \sum_t \sum_u X^2_{\cdot\cdot stu}/QR, & f &= \sum_r \sum_s X^2_{\cdot rs\cdot\cdot}/QTU, \\
g &= \sum_r \sum_t X^2_{\cdot r\cdot t\cdot}/QSU, & h &= \sum_s \sum_t X^2_{\cdot\cdot st\cdot}/QRU, \\
i &= \sum_r \sum_u X^2_{\cdot r\cdot\cdot u}/QST, & j &= \sum_s \sum_u X^2_{\cdot\cdot s\cdot u}/QRT, \\
k &= \sum_t \sum_u X^2_{\cdot\cdot\cdot tu}/QRS, & l &= \sum_r X^2_{\cdot r\cdot\cdot\cdot}/QSTU, \\
m &= \sum_s X^2_{\cdot\cdot s\cdot\cdot}/QRTU, & n &= \sum_t X^2_{\cdot\cdot\cdot t\cdot}/QRSU, \\
o &= \sum_u X^2_{\cdot\cdot\cdot\cdot u}/QRST, & p &= X^2_{\cdot\cdot\cdot\cdot\cdot}/QRSTU, \\
a' &= \sum_q \sum_r \sum_s \sum_t \sum_u X^2_{qrstu}, & b' &= \sum_q \sum_r \sum_s \sum_t X^2_{qrst\cdot}/U, \\
c' &= \sum_q \sum_r \sum_s X^2_{qrs\cdot\cdot}/TU.
\end{aligned}$$

A dot used here in place of a subscript refers to the fact that the sum has been taken over that subscript.

For the comparison of individual means by use of one of the techniques available for this task, e.g., Duncan's multiple range test [11], the following typical variances are given.

a. The difference between two over-all means of dimension u, $\bar{X}_{....u} - \bar{X}_{....u'}$, has a variance which is estimated by

$$\frac{2\text{ M.S. (Error 1)}}{QRST}.$$

b. The difference between two means of dimension t for a given level of dimension u, $\bar{X}_{...tu} - \bar{X}_{...t'u}$, has a variance which is estimated by

$$\frac{2[(U-1)\text{M.S. (Error 1)} + \text{M.S. (Error 2)}]}{QRSU}.$$

The variances of differences between means of dimensions r or s for particular levels of u take analogous forms.

c. The difference between two means of dimension u for a particular level of t, $\bar{X}_{...tu} - \bar{X}_{...tu'}$, has a variance which is estimated by

$$\frac{2\text{ M.S. (Error 1)}}{QRS}.$$

Variances of the differences between two means of dimension u for particular levels of other dimensions take analogous forms.

d. The difference between two means of dimension t, $\bar{X}_{...t.} - \bar{X}_{...t'.}$, has a variance which is estimated by

$$\frac{2\text{ M.S. (Error 2)}}{QRSU}.$$

Variances of the differences between means of dimensions r and s take analogous forms.

e. The difference between two means of dimension t for a particular level of s, $\bar{X}_{..st.} - \bar{X}_{..st'.}$, has a variance which is estimated by

$$\frac{2\text{ M.S. (Error 2)}}{QRU}.$$

Variances of differences between means of dimension t for particular levels of r and also between two means of other combinations of r, s and t follow analogous forms.

A Four-Dimensional System with Levels of One Dimension Correlated ρ_1 and Levels of a Second Dimension Correlated ρ_2

Under different conditions than in the previous section, a change in the assumptions made there on variances and covariances seems warranted. Suppose that Q different subjects are grouped in each of RS subclasses corresponding to dimensions r and s. Now let all QRS subjects be administered U levels of dimension u under all levels of dimension t. Here each subject in the rsth subclass is observed under TU dimension level combinations. Assume that the correlations between observations of levels of dimension u made on the same subject for the same level of t differ from the correlations

between observations taken under different levels of t. For example, the levels of dimension t might refer to various points in time at which each subject is administered all levels of u.

For this situation let the model state that the random variables X_{qrstu} are $QRSTU$-variate normally distributed with

$$(10)\qquad \begin{aligned} &V(X_{qrstu}) = \sigma^2;\\ &\operatorname{cov}(X_{qrstu}, X_{q'r's't'u'})\\ &\qquad = \rho_1\sigma^2, \quad \text{for} \quad q = q', \quad r = r', \quad s = s', \quad t = t', \quad u \neq u',\\ &\qquad = \rho_2\sigma^2, \quad \text{for} \quad q = q', \quad r = r', \quad s = s', \quad t \neq t',\\ &\qquad = 0 \quad \text{otherwise}; \end{aligned}$$

and

$$(11)\qquad E(X_{qrstu}) = \text{the linear function given in (2).}$$

Again transform the variates to independent sets by means of an orthogonal transformation applied twice. First transform X_{qrstu} by means of

$$(12)\qquad \begin{aligned} Y_{qrstw} &= \sqrt{\frac{w}{w+1}}\left\{X_{qrst(w+1)} - \frac{\sum_{u=1}^{w} X_{qrstu}}{w}\right\}, \qquad (w = 1, 2, \cdots, U-1)\cdot\\ Y_{qrstU} &= \sqrt{U}\left\{\frac{\sum_{u=1}^{U} X_{qrstu}}{U}\right\}. \end{aligned}$$

Thereafter, the Y_{qrstU} of (12) are transformed by the same form of the transformation to Z_{qrstU}, of which the Z_{qrsTU} are labeled W_{qrsTU}. Resulting from these transformations is a set (1) of $QRST(U-1)$ variates, Y_{qrstu} ($u = 1, 2, \cdots, U-1$), normally and independently distributed with variance $\sigma_1^2 = \sigma^2\,[1 - \rho_1]$; a set (2) of $QRS(T-1)$ variates, Z_{qrstU} ($t = 1, 2, \cdots, T-1$), normally and independently distributed with variance $\sigma_2^2 = \sigma^2\,[1 + (U-1)\rho_1 - U\rho_2^2]$; and a set (3) of QRS variates, W_{qrsTU}, which are normally and independently distributed with variance $\sigma_3^2 = \sigma^2\,[1 + (U-1)\rho_1 + (T-1)U\rho_2]$. Again these variates are independent between sets.

It follows that the sums of squares associated with set (1), set (2), and set (3) are given by (13), (14), and (15), respectively.

$$(13)\qquad \begin{aligned} &\sum_q \sum_r \sum_s \sum_t \sum_{u=1}^{U-1} [Y_{qrstu} - E(Y_{qrstu})]^2\\ &\quad = \sum_q \sum_r \sum_s \sum_t \sum_u \Big[X_{qrstu} - \frac{X_{qrst\cdot}}{U} - \alpha(4, u) - \beta(4, ru)\\ &\quad - \beta(5, su) - \beta(6, tu) - \gamma(2, rsu) - \gamma(3, rtu) - \gamma(4, stu)\\ &\quad - \delta(rstu)\Big]^2. \end{aligned}$$

$$(14)\quad \sum_q \sum_r \sum_s \sum_{t=1}^{T-1} [Z_{qrstU} - E(Z_{qrstU})]^2$$

$$= U \sum_q \sum_r \sum_s \sum_t \left[\frac{X_{qrst\cdot}}{U} - \frac{X_{qrs\cdot\cdot}}{TU} - \alpha(3, t) - \beta(2, rt) - \beta(3, st) - \gamma(1, rst) \right]^2.$$

$$(15)\quad \sum_q \sum_r \sum_s [W_{qrsTU} - E(W_{qrsTU})]^2$$

$$= TU \sum_q \sum_r \sum_s \left[\frac{X_{qrs\cdot\cdot}}{TU} - \mu - \alpha(1, r) - \alpha(2, s) - \beta(1, rs) \right]^2.$$

As before, tests of hypotheses on parameters involved in the three variate sets may be derived. These tests are given below and the analysis of variance follows in Table 2.

In set (1) the test of the hypothesis of no third-order interaction, H_0 : $\delta(rstu) = 0$, is accomplished by forming the F ratio

$$(16)\quad F(\delta) = \frac{\text{S.S.}[\delta(rstu)]/(R-1)(S-1)(T-1)(U-1)}{\text{S.S.(Error 1)}/(Q-1)(RST)(U-1)}.$$

Other hypotheses in set (1) are tested by forming analogous F ratios.

Again in set (2) we test the hypothesis, $H_0 : \gamma(1, rst) = 0$, by referring the ratio

$$(17)\quad F[\gamma(1)] = \frac{\text{S.S.}[\gamma(1, rst)]/(R-1)(S-1)(T-1)}{\text{S.S.(Error 2)}/(Q-1)(RS)(T-1)}$$

to its proper F distribution and other hypotheses of set (2) are tested by forming comparable F ratios.

Finally, in set (3) the hypothesis, $H_0 : \beta(1, rs) = 0$, can be tested by forming the F ratio

$$(18)\quad F[\beta(1)] = \frac{\text{S.S.}[\beta(1, rs)]/(R-1)(S-1)}{\text{S.S.(Error 3)}/(Q-1)(RS)}.$$

The three remaining hypotheses in set (3) may be tested by forming comparable F ratios. The analysis of variance for this system is given in Table 2.

The notation used in the sum of squares column of Table 2 has been defined in (9), and the expected mean squares due to error are

$$E\{\text{M.S. (Error 1)}\} = \sigma_1^2 = \sigma^2(1 - \rho_1),$$

$$E\{\text{M.S. (Error 2)}\} = \sigma_2^2 = \sigma^2[1 + (U-1)\rho_1 - U\rho_2],$$

$$E\{\text{M.S. (Error 3)}\} = \sigma_3^2 = \sigma^2[1 + (U-1)\rho_1 + (T-1)U\rho_2].$$

TABLE 2

Analysis of Variance for a Four Dimensional Layout in Which the Observations of Levels of One Dimension Are Correlated ρ_1 and of the Levels of Another Dimension Are Correlated ρ_2

Source of Variation	Degrees of Freedom	Sum of Squares
Total Set (1)	QRST(U-1)	a'-b'
(Between "u" within "t" summed over individuals)		
α(4,u)	U-1	o-p
β(4,ru)	(R-1)(U-1)	i-ℓ-o+p
β(5,su)	(S-1)(U-1)	j-m-o+p
β(6,tu)	(T-1)(U-1)	k-n-o+p
γ(2,rsu)	(R-1)(S-1)(U-1)	c-f-i-j+ℓ+m+o-p
γ(3,rtu)	(R-1)(T-1)(U-1)	d-g-i-k+ℓ+n+o-p
γ(4,stu)	(S-1)(T-1)(U-1)	e-h-j-k+m+n+o-p
δ(rstu)	(R-1)(S-1)(T-1)(U-1)	a-b-c-d-e+f+g+h+i+j+k-ℓ-m-n-o+p
Error (1)	(Q-1)(RST)(U-1)	a'-b'-a+b
Total Set (2)	QRS(T-1)	b'-c'
(Between "t" summed over individuals)		
α(3,t)	T-1	n-p
β(2,rt	(R-1)(T-1)	g-ℓ-n+p
β(3,st)	(S-1)(T-1)	h-m-n+p
γ(1,rst)	(R-1)(S-1)(T-1)	b-f-g-h+ℓ+m+n-p
Error (2)	(Q-1)(RS)(T-1)	b'-c'-b+f
Total Set (3)	QRS	c'
(Between individuals)		
μ	1	p
α(1,r)	R-1	ℓ-p
α(2,s)	S-1	m-p
β(1,rs)	(R-1)(S-1)	f-ℓ-m+p
Error(3)	(Q-1)(RS)	c'-f

The variances of certain typical differences in means follow.

a. The difference between two means of dimension *u*, $\bar{X}_{....u} - \bar{X}_{....u'}$, has a variance which is estimated by

$$\frac{2\text{ M.S. (Error 1)}}{QRST}.$$

b. The difference between two means of dimension t, $\bar{X}_{\ldots t.} - \bar{X}_{\ldots t'.}$, has a variance which is estimated by

$$\frac{2\text{ M.S. (Error 2)}}{QRSU}.$$

c. The difference between two means of dimension s, $\bar{X}_{..s..} - \bar{X}_{..s'..}$, has a variance which is estimated by

$$\frac{2\text{ M.S. (Error 3)}}{QRTU}.$$

Variances of the differences between means of r take analogous forms.

d. The difference between two means of dimension u for a particular level of t, $\bar{X}_{\ldots tu} - \bar{X}_{\ldots tu'}$, has a variance which is estimated by

$$\frac{2\text{ M.S. (Error 1)}}{QRS}.$$

Variances of the differences between means of dimension u for particular levels of r and s take analogous forms.

e. The difference between two means of dimension t for a particular level of u, $X_{\ldots tu} - X_{\ldots t'u}$, has a variance which is estimated by

$$\frac{2[(U-1)\text{M.S. (Error 1)} + \text{M.S. (Error 2)}]}{QRSU}.$$

f. The difference between two means of dimension s for a particular level of u, $X_{..s.u} - X_{..s'.u}$, has a variance which is estimated by

$$\frac{2[(U-1)\text{M.S. (Error 1)} + \text{M.S. (Error 3)}]}{QRTU}.$$

Variances of the differences between means of r for a particular level of u take analogous forms.

g. The difference between two means of dimension s for a particular level of r, $X_{.rs..} - X_{.rs'..}$, has a variance which is estimated by

$$\frac{2\text{M.S. (Error 3)}}{QTU}.$$

Variances of the differences between means of r for a particular level of s take analogous forms.

Systems of Lower Dimensionality

Due to limitations on space it will not be possible to examine special cases of the previous systems wherein the number of dimensions is less than

four. It is suggested that systems of lower dimensionality be viewed as cases of the preceding systems in which certain subscripts, say r, s, or t, have been deleted and Tables 1 and 2 be altered to take care of these setups. Thus, all parameters in Table 2 containing a deleted subscript will be omitted from analysis and the sums of squares and degrees of freedom which remain will be altered by striking out either the deleted subscripts or corresponding limits of summation as the case might be. Tests of hypotheses are made in the same manner as before.

Problems in Interpretation

In classification systems, whether the levels of the dimensions are experimental or not, the difficulties encountered in interpreting the results of an analysis usually increase when the number of dimensions are increased. This is particularly true when it is desired to explain the meaning of significant higher-order interactions. Methods useful in the explanation of significant higher-order interactions (see [12]) for systems in which the observations are assumed to be uncorrelated may be used also in the setups presented in this paper.

Normally it is necessary to compare individual means in order to explain both significant main effects and interactions, this being accomplished by tests of significance such as multiple range tests. In the previously presented models, however, this task becomes more complicated since the analyses contain more than one error variance. Although the variances of differences between particular means which previously have been given should facilitate such comparisons, it should be pointed out that the degrees of freedom for variances which involve more than one error mean square may be approximated by means of methods given by Taylor [16].

REFERENCES

[1] Binder, A. The choice of an error term in analysis of variance designs. *Psychometrika*, 1955, 20, 29–50.

[2] Cochran, W. G. and Cox, G. M. *Experimental designs.* (2nd ed.) New York: Wiley, 1957.

[3] Collier, R. O., Jr. Experimental designs in which the observations are assumed to be correlated. Unpublished doctoral dissertation, Univ. Minnesota, 1956.

[4] Edwards, A. L. *Experimental design in psychological research.* New York: Rinehart, 1950.

[5] Halperin, M. Normal regression theory in the presence of intraclass correlation. *Ann. math. Statist.*, 1951, 22, 573–580.

[6] Hartley, H. O. Some recent developments in analysis of variance. *Commun. pure appl. Math.*, 1955, 8, 47–57.

[7] Hoyt, C. J., Stunkard, C. L., *et al.* A comparison of two methods of instruction in beginning drawing. *J. exp. Educ.*, 1952, 20, 265–279.

[8] Kempthorne, O. *The design and analysis of experiments.* New York: Wiley, 1952.

[9] Kogan, L. S. Analysis of variance—repeated measurements. *Psychol. Bull.*, 1948, 45, 131–143.

[10] Kogan, L. S. Variance designs in psychological research. *Psychol. Bull.*, 1953, **50**, 1–40.
[11] Kramer, C. Y. Extension of multiple range tests to group correlated means. *Biometrics*, 1957, **13**, 13–18.
[12] Lindquist, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.
[13] Moonan, W. J. Simultaneous examination and method analysis by variance algebra. *J. exp. Educ.*, 1955, **23**, 253–257.
[14] Moonan, W. J. An analysis of variance method for determining the external and internal consistency of an examination. *J. exp. Educ.*, 1956, **24**, 239–244.
[15] Nandi, H. K. A mathematical set-up leading to analysis of a class of designs. *Sankhyā*, 1947, **8**, 172–176.
[16] Taylor, J. The comparison of pairs of treatments in split-plot experiments. *Biometrika*, 1950, **37**, 443–444.
[17] Wilks, S. S. *Mathematical statistics.* Princeton: Princeton Univ. Press, 1950.
[18] Yates, F. The principles of orthogonality and confounding in replicated experiments. *J. agric. Sci.*, 1933, **23**, 109–145.
[19] Yates, F. Complex experiments. *Suppl. J. roy. statist. Soc.*, 1935, **2**, 181–223.
[20] Yates, F. The design and analysis of factorial experiments. Imp. Bur. soil Sci., Tech. Comm. No. 35, Harpenden, England, 1937.

Manuscript received 5/2/57
Revised manuscript received 1/17/58

PSYCHOMETRIKA—VOL. 23, NO. 3
SEPTEMBER, 1958

THURSTONE'S ANALYTICAL METHOD FOR SIMPLE STRUCTURE AND A MASS MODIFICATION THEREOF*

ROBERT R. SOKAL†
UNIVERSITY OF KANSAS

The analytical method for simple structure proposed by Thurstone is applied to four separate cases and found to yield satisfactory results. The simple structure obtained by Thurstone's method is found to match closely that obtained by other methods and corresponds to the true structure of the matrix in those cases where true structure is known. Difficulties about the choice of the correct trial vector led the writer to develop a modification of Thurstone's method, useful where high speed computational facilities are available. Instructions are given for this so-called mass modification, and the procedure is illustrated with a 5-factor, 14-variable example. While the results do not fully correspond to a previous graphical solution, it can be argued that the results obtained by the new method show an improved simple structure. The modified method is applied to three other correlation matrices, yielding in each case a satisfactory simple structure.

The purpose of this paper is: (1) to give several examples of the application of Thurstone's analytical method for finding simple structure in a factor matrix; (2) to compare the results of Thurstone's solution with those from other analytical solutions as well as with results of trial-and-error graphic rotation; (3) to introduce a mass modification of Thurstone's method, which is believed to be superior where high speed computational devices are available.

The student or investigator entering the field of factor analysis at this time need no longer be dismayed by the prospect of seemingly endless trial-and-error rotation of reference vectors interspersed with hours of tedious desk calculator computation. A number of analytical solutions for simple structure have been proposed by several authors. These methods determine a simple structure constellation by arithmetic rather than geometric operations and appear to have removed the trial-and-error element from the procedure.

*Contribution No. 961 from the Department of Entomology, University of Kansas, Lawrence, Kansas.

†Most of the work on which this paper is based was performed during the summer of 1956 when the author was the holder of an Elizabeth Watkins Faculty Scholarship granted by the Kansas University Endowment Association. The work was carried out at the University of Illinois, where the author was privileged to spend the tenure of this scholarship. The writer is indebted to Professor L. H. Lanier, the Chairman of the Psychology Department, who graciously placed space and equipment at his disposal, and to Professor Raymond B. Cattell for his continuing encouragement of and interest in the author's work. Some of the computations were performed on the ILLIAC digital computer. The many courtesies extended by Professor J. J. Nash, Director of the Digital Computer Laboratory of the University of Illinois, are gratefully acknowledged. The writer is indebted to Mr. John R. Hurley for much assistance during the development of the computational routines and for a critical reading of this paper. Expenses in connection with the work were met by a General Research Grant of the University of Kansas.

Computational labor required for these solutions is generally quite heavy but a parallel development, the advent of high speed electronic computers, is eliminating this factor for those investigators fortunate enough to have access to such a machine. Thus the decision on whether to rotate to simple structure may in future be based mainly on the intrinsic merit of that structure for a particular research problem rather than on whether an investigator has the time or fortitude to undertake a search for simple structure by the traditional methods.

Analytical solutions were published almost simultaneously by a number of authors. Carroll [1] proposed a solution based on the minimization of the sums of the cross products of squares of factor loadings. This method yields orthogonal or oblique factors, as desired. Saunders [8] published a solution based on a maximized kurtosis of factor loadings but with factors restricted to orthogonality. Pinzka and Saunders [7] extended this solution to the oblique factor case (oblimax). Neuhaus and Wrigley [6] developed another method (quartimax), based on maximizing the sum of fourth powers of factor loadings. The equivalence of these methods in the case of *orthogonal* factors has been shown by Ferguson [3] working on an objective definition of parsimony in factor analysis. Tucker's method [13] is a computational routine searching for linear constellations of test vectors.

Thurstone's analytical method [12], for simplicity's sake hereafter referred to as TAM, uses a set of arbitrary weights which adjust trial vectors into positions perpendicular to such hyperplanes as exist in the study. The method is briefly reviewed in the following section. Except for the example, discussed in Thurstone's paper [12], the writer has been unable to find published examples of the application of TAM.* Of the several methods listed above, TAM appeared the simplest from the point of view of using a mixed desk-calculator, digital-computer routine. Before applying TAM to his own data the writer tried it out on several other matrices and compared the results of his efforts with either the known structure of the matrix or with results obtained by other methods, as the data permitted. These findings are presented below in the hope that they may prove of interest.

Several complications arose during the application of Thurstone's original method to some of the matrices; the writer subsequently developed a modification of TAM, which he believes is superior when high speed computational facilities are available (mainly for matrix multiplication and inversion). The method is described in the last section of this paper and four examples of its application to actual cases are given.

*A mimeographed report by Rolf Bargmann, entitled "A comparison of new analytical methods for the determination of simple structure," September 1953, was brought to the notice of the author after this manuscript had been completed and submitted for publication. Bargmann applied TAM to the 21-variable, 8-factor F_0 matrix from p. 402 of Thurstone's *Multiple-Factor Analysis*. A satisfactory simple structure resulted. Several improvements of TAM based on Thurstone's single planing method were suggested by Bargmann in his report.

Brief Review of Thurstone's Analytical Method (TAM)

TAM starts with an unrotated centroid factor matrix F_0 of dimensions $n \times k$, representing n variables and k extracted factors. A trial vector is chosen from one of the variables of the study, and the row of factor loadings of variable i in matrix F_0 is transposed into a column vector J_i of direction numbers. Column vector J_i is then normalized to yield column vector P_i, which changes the direction numbers into direction cosines.

Next the column vector of projections v_{ip} of normal-length test vectors j on unit trial vector P_i are computed,

$$v_{ip} = F_o P_i.$$

A weight w_{ip} is assigned to each v_{ip} according to the table of weights given in Thurstone [12] or, in some cases to be described below, according to a scheme developed separately for each reference vector to be determined. These weights form the principal diagonal of a diagonal matrix W of order n. A weighted factor matrix F_w,

$$F_w = WF_o,$$

is computed and a symmetric matrix A is then obtained,

$$A = F_w' F_o.$$

This symmetric matrix is of order k. The next step is the computation of A^{-1}. This matrix is used to compute column vector U,

$$U = A^{-1} P_i.$$

U is normalized to yield Λ_1, the direction cosines of the first reference vector. The column vector Λ_1 defines the adjusted position of the initial trial reference vector P_i. Loadings of the n variables on the new reference vector $V_{i\Lambda}$ are computed by

$$v_{i\Lambda} = F_o \Lambda_1.$$

A second trial vector is chosen from among variables that are in the hyperplane of the first reference vector, a third from those in the hyperplanes of both the first and the second, and so forth. Occasionally convergence toward the simple structure solution may be slow, and it will be necessary to iterate the procedure using Λ_1 as the new unit trial vector. In practice we start with $V_{i\Lambda}$ in place of V_{ip}, weighting the former, and continuing as before.

As will be seen below a trial vector may at times be adjusted into a reference vector that has already been determined (a disheartening experience in the original TAM since it does not become apparent until the tedious computation of the inverse matrix has been completed). This was one of the causes leading up to the development of the mass modification.

Examples of the Application of Thurstone's Analytical Method

The Cattells' Boys

The first factor matrix tested is an artificial one simulating the results of manipulation tests of seven mechanical puzzles (variables) where success is determined by three factors. It was developed by Cattell and Cattell [2] in connection with their method of parallel proportional profiles. They prepared two similar matrices, one to represent the factor structure of boys, the other that of girls. Only the former was employed in the present study.

The unrotated factor matrix V_{0A} ([2], Table III) was used as the initial F_0 matrix. Variable 7 was chosen as the first trial vector. This yielded the

TABLE 1

Comparison of Simple Structure Solutions for Cattells' Boys: TAM, Graphic Rotation, and Original (Known) Factor Matrix

Reference vector	Matrix / Variable	Original	TAM	Graphic rotation*
I	1	60	64	65
	2	-50	-43	-45
	3	-10	-14	-14
	4	00	-06	00
	5	90	92	92
	6	-60	-58	-65
	7	05	06	-01
II	1	-40	-32	-35
	2	-80	-84	-83
	3	80	79	80
	4	50	42	45
	5	-10	06	01
	6	00	-01	-01
	7	00	10	06
III	1	-10	-08	-06
	2	10	05	11
	3	00	02	-04
	4	-60	-59	-63
	5	00	08	08
	6	70	67	67
	7	80	80	81

* Readers who compare these figures with those in Cattell and Cattell [2] will note that the loadings of variables 2 and 3 for all three reference vectors have been interchanged. This is due to a printer's error in the left half of table VI of that paper. The present version is the correct one according to a personal communication of Prof. R. B. Cattell.

Cattells' reference vector III. Variable 3 had the lowest loading on that reference vector and was used as the second trial vector yielding in turn reference vector II of the Cattells. The final trial vector was clearly variable 5, since it was the only variable located in the hyperplanes of both previous RV's. It yielded the first RV of the previous authors. The weights used were the original ones tabled by Thurstone ([12], Table 1). Since the initial factor matrix was specially designed to represent simple structure and consequently exhibited a wide range in the magnitude of the factor loadings it is not surprising that the original weight scale was adequate.

The result of the above procedure is shown in Table 1. The graphic solution and TAM are almost identical and both closely resemble the original structure of the factor matrix. Correlations between the TAM reference vectors are very low ($r_{I.II} = .136$, $r_{I.III} = .096$, $r_{II.III} = .061$) returning almost completely to the initial orthogonal position. In this particular instance TAM appears to give entirely satisfactory results.

The Johnson-Reynolds Data

The next example is a small matrix which had been used previously by two other workers in demonstrating their own analytical solutions for simple structure (Carroll [1], Pinzka and Saunders [7]). The matrix had been originally computed from ten verbal tests administered to 113 students by Johnson and Reynolds [5] and contains 10 variables (the tests) and two factors.

TAM was applied to the factors using variable 5 as a first trial vector. This resulted in a factor similar to factor I of Carroll and Saunders. Since the hyperplane was determined by only one variable the problem was iterated using the $V_{i\Lambda}$ column as a new V_{ip} column. The second factor was found via variable 2 which lay in the hyperplane of factor I. Thurstone's weights ([12], Table 1) had to be modified somewhat since their original magnitudes would not have given enough differentiation. Thurstone's suggestion ([12], p. 180) that the total range of v_{ip} values be divided into a number of equal intervals and that these intervals be assigned values of from 0 for the highest absolute loading to 6 for the lowest such loading was found to be quite effective. In cases such as Wright's chicken bones discussed below, the v_{ip} range was not such as to warrant division into seven classes, and weight classes of from 0 to 4 were assigned this and similar cases.

Table 2 compares the findings by TAM with those of Carroll and Saunders [1, 7]. It can be seen that the solution by TAM is quite similar to that by the previous workers. The iteration of factor I brought the loadings for that factor considerably closer to the loadings found by the other two methods. Unless one can agree on a common optimal criterion of simple structure it is impossible to compare and evaluate the three solutions. Pinzka and Saunders [7] show that their solution is the better by their criterion while Carroll's is the better by his criterion. The f value (Pinzka and Saunders'

TABLE 2

Comparison of Three Analytical Methods for Simple Structure Applied to Johnson-Reynolds Data

Reference vector	Method / Variable	Carroll	Pinzka and Saunders	TAM (1st iteration)	TAM (2nd iteration)
I	1	-232	-242	-119	-224
	2	-099	-108	007	-091
	3	064	056	149	070
	4	315	303	440	325
	5	638	627	739	646
	6	608	597	712	616
	7	629	620	715	636
	8	574	568	638	580
	9	578	568	681	587
	10	281	267	426	292
II	1	538	532	540	not performed
	2	436	427	439	
	3	265	254	269	
	4	256	234	265	
	5	-017	-044	-006	
	6	010	-017	021	
	7	-066	-092	-056	
	8	-113	-136	-105	
	9	021	-005	032	
	10	350	327	359	

criterion) for the TAM solution is .0495, the worst of the three. On the other hand the correlation between the reference vectors is − .7058 for Pinzka and Saunders' solution while it is only − .4862 for TAM. Without prejudging the other analytical solutions it may be said that TAM did provide a satisfactory solution in this instance.

Wright's Chicken Bones

This is a biological correlation matrix based on six measurements of the bones of 276 chickens. These data were first published by Wright [14], who analyzed them by a method analogous to factor analysis. His findings were revised in a later paper [15]. The correlations appear to be accounted for by three common factors [10, 14, 15]. These data have been rotated to simple structure from a centroid factor matrix twice independently by the writer and his associates. The first two columns of Table 3 show these independent solutions. A uniform numbering system has been adopted for the reference vectors. Reference vector I appears to represent a head factor, while reference vectors II and III clearly represent wing and leg factors, respectively. From Table 4 which shows the correlations between reference vectors we note that the wing and leg factors are almost collinear.

The analysis by TAM presented some difficulty. The use of Thurstone's original weights with variable U resulted in an essentially unrotated but reflected factor III rather than in factor II as expected. This difficulty could be overcome by recoding the weights based on the range of factor loadings as discussed in the previous section. An alternative procedure in this six-variable case was simply to array the v_{ip} values by order of absolute magnitude, assign weight 0 to the largest value and 1 to 5 to values of decreasing magnitude.

When these revised weighting techniques were applied to the V_{ip} column vector obtained from variable U, the resultant A matrix could not be inverted during a first attempt on a desk calculator, using the cutomary Dwyer technique and carrying four decimal places. Investigation disclosed that the determinant of the A matrix was close to zero ($|A| = .013$), apparently due to the high correlation between the wing and leg factors. When the inversion was repeated on the ILLIAC in 12-decimal arithmetic, a solution was found which finally yielded reference vector II of the graphic solutions.

The next variable chosen was F, which is in the hyperplane of reference vector II. It yielded previous reference vector III. Again the determinant of the A matrix was very small requiring five-decimal arithmetic. The final trial vector was based on variable B which had been in the hyperplane of both reference vectors found so far. This yielded reference vector I of the previous analyses. In view of the low correlation of this reference vector with the other two, the determinant of the A matrix was considerably larger than zero, and its inversion proceeded without difficulty.

A glance at Table 3 will convince the reader of the similarity of TAM with the two graphic solutions. Table 4 shows the similarity of the three solutions in terms of the correlations between pairs of reference vectors. In spite of initial difficulties TAM again produced essentially the same simple structure as the graphic method.

TABLE 3

Comparison of Simple Structure Solutions for Wright's Chicken Bone Data: Two Graphic Solutions and TAM

Reference vector	Method / Variable	Graphic #1	Graphic #2	TAM
I	L	49	47	48
	B	50	54	49
	H	09	05	08
	U	02	02	00
	F	01	03	-01
	T	06	08	04
II	L	01	09	03
	B	-04	-01	-02
	H	27	32	29
	U	33	33	35
	F	-01	01	01
	T	-02	01	00
III	L	-04	00	-02
	B	00	04	01
	H	-01	10	02
	U	-04	10	-01
	F	30	41	33
	T	29	40	32

Explanation of variables: L = head length, B = head breadth, H = humerus length, U = ulna length, F = femur length, T = tibia length.

TABLE 4

Correlations between Reference Vectors of Simple Structure Solutions of Table 3

Method	Correlation coefficients		
	I·II	I·III	II·III
Graphic #1	-.22	-.21	-.86
Graphic #2	-.22	-.18	-.82
TAM	-.21	-.21	-.85

Stroud's Termite Soldiers

These data, published by Stroud [11], are based on a correlation matrix involving 14 physical measurements of soldiers of 48 species of the termite genus *Kalotermes*. The correlations were factored by the centroid method; the unrotated F_0 matrix is given in Table II of Stroud's paper. Five factors, extracted from the correlations, were rotated by a trial-and-error graphic method. The rotated solution is given in Table V of Stroud [11]; the plots of pairs of reference vectors are in Figure 3 of the same article.

The writer obtained only one reference vector from these data by TAM. During the analysis of this example the idea of a mass modification came to the writer, and it was decided to carry out the mass analysis of Stroud's data reported in this paper. The single trial vector used was based on variable 12, following Thurstone's [12] instructions that the first trial vector should be based on a test (variable) which has some relatively high correlations in the correlation matrix and which also has an appreciable number of relatively low coefficients. However, the reference vector obtained showed only a poorly defined hyperplane and did not match any of the reference vectors found by Stroud. This would appear to be the first failure of TAM to yield a reference vector corresponding to one obtained graphically during a first trial. A discussion of this case will be deferred until the discussion of the analysis of the Stroud data by the mass modification of TAM.

A Mass Modification of Thurstone's Analytical Method

Description of the Procedure

The mass modification of TAM was developed in response to several problems which arose during the application of this method to the matrices of the preceding section. Some of these problems were the following.

(1) Occasionally a given trial vector would not yield a reference vector with a well-defined hyperplane.

(2) At other times a given variable would result in a well-defined reference vector, but a second variable would have given an even clearer simple structure. It might be argued in such a case that iteration of the reference vector from the first variable would also have yielded better definition. However, it is of course desirable to limit the number of iterations to an essential minimum.

(3) At times the prescribed procedure for finding new trial vectors in TAM produced two collinear reference vectors and lost one of the significant reference vectors in the study. This particular topic is discussed further in Stroud's termite soldier example below.

The essentials of the mass modification of TAM, which will subsequently be referred to as MTAM, are the simultaneous use of every variable as a test

vector. Thus from an $n \times k$ unrotated factor matrix, F_0, where n is the number of variables and k is the number of factors, n trial vectors and consequently n reference vectors are computed, which however represent only k dimensions. The advantages of this method are as follows.

(1) A uniform treatment of every factor matrix without specific decisions to be made before each reference vector is established.

(2) The best definition for a given reference vector can be chosen from among those provided by several trial vectors.

(3) A choice can be made about the obliqueness of reference vectors since a Λ matrix for all n reference vectors will be obtained.

The advantages of MTAM are, however, predicated upon the availability of high speed computing machines since they involve for each iteration the computation of n inverse matrices of order k.

The computational procedure for MTAM is given below.

(1) The procedure starts with the unrotated factor matrix F_0 of order $n \times k$.

(2) F_0 is normalized by rows to F_N. This gives all possible unit trial vectors P_i.

(3) Compute

$$V_{ip} = F_o F_N',$$

where V_{ip} is of order $n \times n$. This yields all possible column vectors v_{ip}. Each column of V_{ip} is a column vector v_{ip}.

(4) Weights are assigned separately to each v_{ip} based on the magnitude and range of the v_{ip} values as described above. A routine procedure which can be given to a clerk or built into a computer routine can be easily set up. The weights are set up as $n \times n$ diagonal matrices W_1 through W_n.

(5) Matrices W_1 through W_n are postmultiplied by F_0. There are, of course, n such multiplications, which yield n matrices (of order $n \times k$) F_{W1} through F_{Wn}. In general

$$F_{Wi} = W_i F_o.$$

(6) Premultiply F_{Wi} by constant matrix F_0' to give n symmetric matrices of type A_i (of order $k \times k$).

$$A_i = F_o' F_{wi}.$$

(7) Compute inverses A_i^{-1} for n matrices A_i.

(8) Split F_N' into n column vectors of type F_{Ni}'.

Then postmultiply the n inverse matrices by the corresponding column vectors F_{Ni}' to obtain n column vectors U_i,

$$U_i = A_i^{-1} F_{Ni}'.$$

(Some workers with TAM have preferred to solve the simultaneous equation $AU = F'_N$ for U instead of the two-step procedure originally formulated by Thurstone and followed here. The choice is entirely one of convenience and availability of programs for a given computer.)

(9) Assemble the n column vectors U_i into a single matrix U and normalize by columns to matrix Λ (of order $k \times n$).

(10) The final step consists of computing the $n \times n$ reference vector matrix $V_{j\Lambda}$,

$$V_{j\Lambda} = F_o\Lambda.$$

The writer performed these procedures step by step simultaneously for seven different factor matrices (in about 3 hours computer time plus 3 days handling time, discounting errors made as the technique was being perfected). However, it would seem preferable to program the entire sequence as a single routine. This would then permit iterations. Machine output should include $V_{j\Lambda}$ and Λ (or preferably $\Lambda'\Lambda$). In the case of matrices of appreciable size, such as 50 tests (variables) and 10 factors, storage facilities on most computers would be inadequate to handle all the steps in a single procedure. In such a case the $n \times n$ weight matrices W_1 through W_n would have to be entered singly into the machine and steps (5) through (8) would have to be performed in sequence separately for each trial vector. This would also permit dropping out any trial vector if the operator so desired. Subsequent procedures for finding optimal simple structure are discussed in the next section in relation to a concrete example.

The Mass Modification Applied to Stroud's Data

The 14-variable, 5-factor matrix of termite soldier measurements by Stroud [11] mentioned previously was subjected to the MTAM procedure just described. The matrix was completely processed once as an experimental problem to study the feasibility of MTAM. After the method had been found to be effective new weights were assigned to the resulting $V_{j\Lambda}$ matrix and the problem was iterated. Table 5 shows $V_{j\Lambda}$ of this second iteration. It gives the factor loadings (correlations with the reference vectors) based on the fourteen trial vectors corresponding to the fourteen variables. Table 6 shows the correlations between these fourteen reference vectors calculated by

$$C_R = \Lambda'\Lambda.$$

The incumbent problem is to select from the fourteen reference vectors five which will yield a satisfactory simple structure. At the end of an MTAM analysis one is, of course, able to carry out very simply and elegantly Thurstone's [12] original procedure of starting out with a likely trial vector; then using a variable in the hyperplane of the first, followed by a variable in the hyperplane of both reference vectors, and so forth. All that is needed

TABLE 5

$V_{j\Lambda}$ Matrix (Correlations of Variables with Reference Vectors) after Second MTAM Iteration to Stroud's Centroid Factor Matrix for Termite Soldiers

Variables	Reference vectors													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	32	11	07	21	-10	-12	11	02	07	17	09	04	26	28
2	14	57	52	08	-11	-05	04	02	-05	04	-08	16	07	-02
3	00	62	64	-06	08	15	-08	00	05	-03	05	-12	-10	06
4	23	01	-01	33	20	16	-17	-08	-10	32	-08	08	31	09
5	-07	-10	-05	-10	45	42	-26	06	07	18	12	-01	04	-01
6	-07	14	17	14	47	47	-34	-01	-04	21	-01	02	05	-10
7	07	00	-01	-04	-07	-08	23	30	34	-06	36	17	02	16
8	-07	06	04	-06	03	04	15	36	29	-09	29	36	-04	-08
9	05	-01	02	-06	01	00	18	26	37	-06	41	01	-02	25
10	21	00	-01	-30	22	18	-17	-06	-07	31	-04	04	28	11
11	07	-03	00	01	11	09	06	17	27	04	32	-04	04	24
12	08	-01	-08	18	03	01	04	18	02	12	00	49	19	-18
13	23	-10	-12	25	10	05	-03	02	04	24	07	08	27	18
14	24	02	03	10	-03	-05	07	-03	13	11	18	-24	14	40
Number of variables in hyperplane	8	8	10	6	6	8	6	9	8	6	8	8	7	5

is to look up the values of $v_{j\Lambda}$ in $V_{j\Lambda}$. However, the original TAM which proceeds somewhat blindly from reference vector to reference vector will soon fail to appeal to the experimenter. A multiplicity of possible RV choices becomes available. Some result in five reference vectors, others only in four. Thus the outcome of TAM is to some degree dependent upon the choice of an initial variable. Having gone through the MTAM computational procedures the experimenter has at his disposal considerable added information, including the number of variables in the hyperplanes of all possible reference vectors as well as all mutual correlations between these vectors. It should

TABLE 6

C_R Matrix (Correlations between Reference Vectors) of Table 5

Variables	Reference vectors													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	1.00													
2	-.39	1.00												
3	-.49	.98	1.00											
4	.82	-.36	-.44	1.00										
5	-.31	-.04	.06	.20	1.00									
6	-.47	.13	.24	.05	.98	1.00								
7	-.04	.01	-.04	-.54	-.90	-.84	1.00							
8	-.71	.24	.27	-.90	-.35	-.22	.69	1.00						
9	-.55	.06	.12	-.88	-.40	-.30	.75	.91	1.00					
10	.73	-.37	-.41	.98	.40	.24	-.70	-.94	-.90	1.00				
11	-.51	-.03	.05	-.84	-.36	-.30	.72	.85	.99	-.84	1.00			
12	-.42	.33	.23	-.36	-.29	-.19	.39	.64	.28	-.48	.18	1.00		
13	.93	-.45	-.55	.96	-.03	-.20	-.31	-.82	-.76	.90	-.71	-.34	1.00	
14	.74	-.51	-.50	.37	-.34	-.48	.22	-.40	-.02	.34	.07	-.72	.53	1.00

be possible from this information to develop a procedure leading to a unique and optimal set of reference vectors.

Considerable time and effort have been spent by the writer on developing a simple "cookbook" procedure for obtaining an optimal simple structure using $V_{f\Lambda}$ and C_R of MTAM. Regrettably, no such simple routine valid for all investigated examples could be devised. However, a series of steps may be outlined which in the cases tested has yielded satisfactory results.

(1) Examine C_R (and also check $V_{f\Lambda}$ and Λ) for pairs or groups of highly correlated ($|r| > .8$) reference vectors. These are usually based on closely related variables (in psychological matrices tests measuring nearly identical psychological functions). Some method of cluster analysis of C_R is helpful in this procedure. In the matrix of Table 6 the following nonoverlapping clusters could be isolated considering only correlation coefficients larger than 0.8: 2–3, 5–6–7; overlapping clusters: 1–4–13, 4–8–9–10–11, 8–13, 10–13. The variables in a given cluster usually represent a certain reference vector. The choice of which variable is to represent the factor is made at a later stage, except in cases such as those described below where the reference vectors in a cluster are essentially collinear (e.g., reference vectors O and P in the housefly data were correlated at $r = .999$). In such cases it is, of course, immaterial which reference vector is used to represent the cluster.

(2) Count the low correlations ($|r| < .2$) of each reference vector with other reference vectors (see Table 6). Below is shown the code number of each reference vector followed by the number of vectors with which it shows low correlations:

1–1	8–0
2–5	9–3
3–4	10–0
4–1	11–4
5–3	12–2
6–3	13–1
7–3	14–2

In some cases, such as Wright's chicken bones discussed above, correlations as low as .2 do not exist. In such circumstances the definition of a low correlation must be altered to permit the formation of groups of relatively unrelated reference vectors.

(3) Count the number of variables in the hyperplane of each reference vector (factor loadings between $-.10$ and $+.10$.) This has been done for the Stroud matrix at the foot of Table 5.

Now proceed to select the reference vectors for a simple structure in this matrix. Our eventual criteria for simple structure will be the five requisites listed by Fruchter ([4], p. 110), namely:

(*a*) Each variable should have at least one loading close to zero.

(*b*) There should be, for each factor column, at least as many tests with zero loadings as there are factors.

(*c*) For every pair of factors there should be several variables with projections on one factor vector but not on the other.

(*d*) For problems having four or more factors, a large proportion of the variables should have negligible loadings on any pair of factors.

(*e*) Only a small number of variables should have appreciable loadings on any pair of factors.

However, attention is first directed to securing a large number of zero or near zero loadings for each reference vector adopted and low correlations between reference vectors whenever possible.

Array all of the reference vectors in accordance with their suitability with respect to the number of variables in the hyperplanes of the reference vectors: 3, 8, 1–2–6–9–11–12, 13, 4–5–7–10, 14 (cf. bottom line of Table 5). When examined for the largest number of low correlations with other reference vectors the following array by suitability resulted: 2, 3–11, 5–6–7–9, 12–14, 1–4, 13, 8–10 (cf. Table 6).

Checking back with the high correlation clusters of reference vectors, only one representative each of the following two clusters should be chosen: 2–3 and 5–6–7. Reference vector 2 was chosen over 3 arbitrarily since the two are almost collinear. Reference vector 6 seems slightly more suitable to represent 5–6–7 since it has two more variables in its hyperplane than 5 or 7. Reference vector 1 was chosen to represent 1–4–13 by similar reasoning. In cluster 4–8–9–10–11 the choice fell on 11 as the best representative by both criteria rather than 8 which, while better by the first array, was much worse by the second. Variables in the remaining overlapping clusters (8–13, 10–13) having been represented already, the fifth reference vector was chosen from variables 12 and 14 which had not entered into any cluster. These two reference vectors were quite highly correlated with each other ($r = -.72$); 12 was chosen over 14 by referring to the suitability of arrays. It is admitted that this procedure lacks a unique solution, but the subjective decisions required are usually between quite similar alternatives. Thus the result would not have been appreciably different if reference vector 5 had taken the place of 6, 3 had been in place of 2, 8 in place of 11, etc.

Table 7 shows the simple structure reference vector matrix adopted by graphic methods. Only reference vectors 2 and 6 are clearly identical to Stroud's E and D, respectively. Reference vector 12 is a poor representative of Stroud's B, while vectors 1 and 11 together determine the covariance accounted for by Stroud's A and C. The question may therefore be raised whether MTAM has achieved its goal in view of discrepancies with Stroud's solution. However, the reader may convince himself by inspecting Table 7 that the MTAM rotated reference vector matrix shows excellent simple structure. Plots of the reference axes against each other (not shown) confirm

TABLE 7

Stroud's Termite Soldiers: MTAM Solution compared with Stroud's Graphic Solution

Variables	MTAM Reference vectors (1)	(2)	(6)	(11)	(12)
1	32	11	-12	09	04
2	14	57	-05	-08	16
3	00	62	15	05	-12
4	23	01	16	-08	08
5	-07	-10	42	12	-01
6	-07	14	47	-01	02
7	07	00	-08	36	17
8	-07	06	04	29	36
9	05	-01	00	41	01
10	21	01	18	-04	04
11	07	-03	09	32	-04
12	08	-01	01	00	49
13	23	-10	05	07	08
14	24	02	-05	18	-24

Variables	Stroud's graphic solution Reference vectors A	B	C	D	E
1	36	01	30	07	10
2	00	06	26	-06	57
3	00	-04	-01	09	52
4	19	-05	32	22	05
5	04	03	-03	40	-12
6	-06	-01	04	46	12
7	25	26	04	-11	-04
8	03	38	02	-01	-06
9	30	18	-06	-04	-10
10	20	-05	28	24	02
11	28	10	-02	-07	-11
12	-01	28	30	04	09
13	28	03	25	10	-08
14	40	-11	08	-03	-07

these conclusions. The MTAM solution has a greater number of variables in the hyperplanes and the reference vectors are slightly less correlated (see Table 8) than they are by Stroud's findings. Although this is not the place to discuss the findings in detail it may be added that the new simple structure is somewhat more interpretable biologically speaking.

Finally it may be appropriate to mention a supposition of the author which he has so far been unable to test. If each matrix were iterated a number of times until stability of $V_{i\lambda}$ were reached, it may well be that the n reference vectors would have collapsed into a number of positions corresponding

TABLE 8

Correlations between Reference Vectors of Table 7

	MTAM Reference vectors				
	1	2	6	11	12
1	1.00				
2	-.39	1.00			
6	-.47	.13	1.00		
11	-.51	-.03	-.28	1.00	
12	-.42	.33	-.18	.18	1.00

	Stroud's graphic solution Reference vectors				
	A	B	C	D	E
A	1.00				
B	-.47	1.00			
C	.43	-.53	1.00		
D	-.38	-.49	.01	1.00	
E	-.62	.26	.05	.03	1.00

to the number, k, of factors in the matrix. In such a case the choice of the correct reference vectors would, of course, not present any problem and advantages (2) and (3) of MTAM given above would vanish. Until MTAM has been programmed on a computer (to permit rapid iteration) this supposition cannot be tested. The single iteration performed in connection with the MTAM study of the Stroud data yielded neither conclusive nor even suggestive evidence on this point.

The Application of MTAM to Three Other Matrices

Graphic solutions for simple structure are unavailable for these matrices, so that the success of MTAM has to be judged by the criteria for simple structure mentioned above. The biological plausibility of the structure obtained is not discussed, being beyond the stated purpose of this paper. These matrices and others will be discussed in detail in several publications in preparation.

TABLE 9

MTAM Solution for Correlation Matrix of 14 Morphological Characters in Houseflies. Simple Structure Matrix Above, Correlations between Reference Vectors Below

Variables	Reference vectors					
	I	II	III	IV	V	VI
A	34	05	19	-03	32	04
B	46	12	00	01	25	-05
C	48	-03	41	00	-09	01
D	44	17	19	-01	-03	03
E	12	30	17	03	-03	-01
F	02	45	-03	-04	-05	-02
G	-01	00	40	00	08	-05
J	-04	03	-04	61	-02	05
K	03	-04	02	60	00	-05
L	-04	08	22	-01	26	05
M	09	07	16	10	27	04
N	06	-07	17	-04	44	-03
O	05	-05	-01	01	02	77
P	-04	04	00	-01	-03	78

Reference vectors	Reference vectors I	II	III	IV	V	VI
I	1.00					
II	-.32	1.00				
III	.03	-.67	1.00			
IV	.05	-.06	.02	1.00		
V	.02	-.39	.12	-.02	1.00	
VI	-.15	.02	-.05	-.17	-.13	1.00

Table 9 is based on an MTAM analysis of an orthogonal factor matrix of six factors and fourteen morphological characters in 421 houseflies, *Musca*

TABLE 10

MTAM Solution for Correlation Matrix of 18 Morphological Characters in the Aphid, Pemphigus populi-transversus. Simple Structure Matrix Above, Correlations between Reference Vectors Below

Variables	Reference vectors						
	I	II	III	IV	V	VI	VII
B	25	28	04	13	04	-02	-06
C	42	02	07	04	-05	-01	-17
D	-01	53	06	08	-11	-04	10
E	37	34	-06	07	04	02	04
F	17	46	-03	-04	-01	-04	-01
G	28	15	05	27	05	-07	-05
H	06	28	06	16	13	08	-06
I	09	02	-02	34	-01	-11	02
L	02	-04	43	07	37	-01	00
M	08	04	54	01	09	-08	-07
N	03	-03	21	34	-08	-01	11
O	02	-03	-04	48	-05	-07	40
P	-06	-01	08	-09	43	02	04
Q	06	00	21	-05	17	07	12
R	-10	01	05	10	-15	32	04
S	08	-01	-09	-02	07	52	-02
T	-12	00	-03	07	20	-08	14
U	-19	07	-06	08	14	-01	32

Reference vectors	Reference vectors						
	I	II	III	IV	V	VI	VII
I	1.00						
II	-.22	1.00					
III	-.19	-.04	1.00				
IV	.12	-.66	-.27	1.00			
V	-.18	-.02	.10	-.22	1.00		
VI	.14	.00	-.26	.14	-.15	1.00	
VII	-.56	.13	-.12	.30	.23	-.13	1.00

domestica. The solution appears to be a satisfactory simple structure as seen from the factor matrix in the upper part of the table and the correlation matrix between reference vectors in the lower part. The simple structure is interpretable from the biological point of view.

Table 10 shows the result of MTAM on an orthogonal factor matrix of seven factors and eighteen morphological characters in 343 aphids of the species *Pemphigus populi-transversus.* The original correlations were product moment coefficients based on covariances within clones (galls). Again the solution appears to be a satisfactory simple structure both from the statistical and biological points of view.

TABLE 11

MTAM Solution for Q-Type Correlation Matrix of 4 Genera of Bees in the Tribe Osmiini. Simple Structure Matrix Above, Correlations between Reference Vectors Below

Variables (species code numbers)	Reference vectors I	II	III	IV
4	64	06	00	-15
5	66	-06	-03	09
8	62	-02	01	12
26	07	22	-09	45
35	-03	75	-02	12
36	00	82	05	-07
40	-06	-04	-02	70
50	08	-03	26	41
67	-01	01	88	-01
68	-01	01	88	-02

Reference vectors	Reference vectors I	II	III	IV
I	1.00			
II	-.32	1.00		
III	-.34	-.30	1.00	
IV	-.33	-.01	-.10	1.00

A final example is shown in Table 11. This is a matrix based on correlations between ten species representing four genera of solitary bees of the tribe Osmiini. The Q correlations were based on measurements of 122 variables. The simple structure obtained by MTAM is very evident. The reference vectors represent genera of these bees and the correlations between these genera. The primary loadings and correlations between primary factors based on the MTAM solution of this matrix have been published elsewhere [9].

If we may generalize from the four matrices described above, MTAM appears to provide entirely satisfactory simple structure solutions with a minimum of (subjective) decisions required from the investigator.

REFERENCES

[1] Carroll, J. B. An analytical solution for approximating simple structure in factor analysis. *Psychometrika*, 1953, **18**, 23–38.

[2] Cattell, R. B. and Cattell, A. K. S. Factor rotation for proportional profiles: analytical solution and an example. *Brit. J. statist. Psychol.*, 1955, **8**, 81–91.

[3] Ferguson, G. A. The concept of parsimony in factor analysis. *Psychometrika*, 1954, **19**, 281–290.

[4] Fruchter, B. *Factor analysis*. New York: Van Nostrand, 1954.

[5] Johnson, D. M. and Reynolds, F. A factor analysis of verbal ability. *Psychol. Rec.*, 1941, **4**, 183–195.

[6] Neuhaus, J. O. and Wrigley, C. The quartimax method: an analytical approach to orthogonal simple structure, *Brit. J. statist. Psychol.*, 1954, **7**, 88–92.

[7] Pinzka, C. and Saunders, D. R. Analytic rotation to simple structure, II: Extension to an oblique solution. Educational Testing Service Bulletin, RB–54–31, 1954. (Multilithed)

[8] Saunders, D. R. An analytical method for rotation to orthogonal simple structure. Educational Testing Service Bulletin, RB–53–10, 1953. (Multilithed), (also *Amer. Psychologist*, 1953, **8**, 428. (Abstract))

[9] Sokal, R. R. Quantification of systematic relationships and of phylogenetic trends. Montreal: Proc. Tenth Internat. Congr. Entomology, 1958, in press.

[10] Sokal, R. R. A comparison of five tests for completeness of factor extraction. 1958, (in preparation).

[11] Stroud, C. P. An application of factor analysis to the systematics of *Kalotermes*. *Syst. Zool.*, 1953, **2**, 75–92.

[12] Thurstone, L. L. An analytical method for simple structure. *Psychometrika*, 1954, **19**, 173–194.

[13] Tucker, L. R. The objective definition of simple structure in linear factor analysis. *Psychometrika*, 1955, **20**, 209–225.

[14] Wright, S. General, group, and special size factors. *Genetics*, 1932, **17**, 603–619.

[15] Wright, S. The interpretation of multivariate systems. *Statistics and mathematics in biology*. Ames, Iowa: Iowa State College Press, 1954.

Manuscript received 7/29/57
Revised manuscript received 12/24/57

PSYCHOMETRIKA—VOL. 23, NO. 2
JUNE, 1958

ATTENUATION AND INTERACTION

QUINN MCNEMAR

STANFORD UNIVERSITY

A significance test is proposed for determining whether a correlation coefficient is less than unity by an amount greater than that attributable to errors of measurement.

It is the purpose of this note to point out a connection between correction for attenuation and interaction in the analysis of variance. Let X_{itm} stand for the mth parallel measure of the ith individual on the tth test, with $i = 1, \cdots, N; t = 1, 2;$ and $m = a, b$. In order to give the problem an analysis of variance setting it is necessary to have all scores in comparable form. A common metric will result if, by using the respective means and sigmas, all four sets of scores are separately transformed to standard scores, with mean

TABLE 1

Score Schema

	Test					
	1			2		
	Measure			Measure		
Individual	a (u)	b (v)	Mean (w)	a (x)	b (y)	Mean (z)
1	X_{11a}	X_{11b}	$\bar{X}_{11.}$	X_{12a}	X_{12b}	$\bar{X}_{12.}$
2	X_{21a}	X_{21b}	$\bar{X}_{21.}$	X_{22a}	X_{22b}	$\bar{X}_{22.}$
$\vdots$						
i	X_{i1a}	X_{i1b}	$\bar{X}_{i1.}$	X_{i2a}	X_{i2b}	$\bar{X}_{i2.}$
$\vdots$						
N	X_{N1a}	X_{N1b}	$\bar{X}_{N1.}$	X_{N2a}	X_{N2b}	$\bar{X}_{N2.}$
Mean	$\bar{X}_{.1a}$ =	$\bar{X}_{.1b}$ =	$\bar{X}_{.1.}$ =	$\bar{X}_{.2a}$ =	$\bar{X}_{.2b}$ =	$\bar{X}_{.2.} = 0$

$\bar{X}_{it.} \neq 0, \quad \bar{X}_{i.m} \neq 0, \quad \bar{X}_{i..} \neq 0; \quad \bar{X}_{..a} = \bar{X}_{..b} = \bar{X}_{...} = 0$

of zero and variance of unity. Thus regard X_{itm} as a standard score. These necessary transformations, which will not lead to any loss in generality, will permit certain simplifications in the sequel.

The $4N$ scores, along with possible means, may be set forth as in Table 1, from which it is readily seen that the setup corresponds to a three-way analysis of variance. Keeping in mind that with standard scores certain means (as indicated in Table 1) will be zero, we may write out the analysis of variance depicted in Table 2. The degrees of freedom add to $4N - 4$; that this is correct may be inferred from the fact that for each of the four

TABLE 2

Analysis of Variance

Source	Sum of squares	d.f.	Var. est.
Individuals	$4\sum_{i}(\bar{X}_{i..})^2$	N-1	s^2_i
Tests	Vanishes	0	--
Measures	Vanishes	0	--
T x M	Vanishes	0	--
I x T	$2\sum_{i}\sum_{t}(\bar{X}_{it.}-\bar{X}_{i..})^2$	N-1	s^2_{it}
I x M	$2\sum_{i}\sum_{m}(\bar{X}_{i.m}-\bar{X}_{i..})^2$	N-1	s^2_{im}
I x T x M	$\sum_{i}\sum_{t}\sum_{m}(X_{itm}-\bar{X}_{it.}-\bar{X}_{i.m}+\bar{X}_{i..})^2$	N-1	s^2_{itm}
Total	$\sum_{i}\sum_{t}\sum_{m}(X_{itm})^2$	4N-4	

sets of scores the mean is forced to be zero, hence there are four *linear* restrictions on the deviations entering into the total sum of squares.

Now if true scores were available on the two tests and if the true scores were perfectly correlated, the individual by test interaction would be exactly zero. With fallible scores the correlation would be attenuated; that is, the correlation between $\bar{X}_{i1.}$ and $\bar{X}_{i2.}$ would be less than unity by an amount attributable to error (of measurement), hence the $I \times T$ interaction would not be zero. Significant $I \times T$ interaction, when tested against the $I \times T \times M$ interaction as error, would mean that the correlation corrected for attenuation is significantly less than unity, whereas insignificant $I \times T$ interaction would lead to the acceptance of the hypothesis that the two tests measure identical functions.

For the tedium of converting to standard scores and doing the required computations for an analysis of variance, the tedium of computing correlation coefficients, some of which would ordinarily be needed anyway, may be substituted. The a and b measures on each test may be regarded as scores on parallel halves of the tests, hence each r_{ab} would be a split half reliability coefficient (not stepped up), and the correlation between the averages, $\bar{X}_{i1.}$ and $\bar{X}_{i2.}$, will be precisely the correlation between the underlying sums (of two scores based on halves), i.e., the correlation between the two tests.

Let us now consider the three-way interaction. It can be shown ([4], pp. 296-297) that the $I \times T \times M$ interaction sum of squares reduces to

$$\tag{1} \frac{1}{4} \sum_i [(X_{i1a} - X_{i1b}) - (X_{i2a} - X_{i2b})]^2,$$

which is a function of the difference between parallel measures on test 1 and on test 2. Cumbersome subscripts may be avoided by setting

$$X_{i1a} = u, \qquad X_{i1b} = v, \qquad \bar{X}_{i1.} = w,$$

$$X_{i2a} = x, \qquad X_{i2b} = y, \qquad \bar{X}_{i2.} = z.$$

In what follows it must be remembered that u, v, x and y are in standard score form; w and z are not standard scores.

Expression (1) may be written as

$$\tag{2} \tfrac{1}{4}[\sum (u - v)^2 + \sum (x - y)^2 - 2 \sum (u - v)(x - y)],$$

which becomes

$$\begin{aligned} &\tfrac{1}{4}(N\sigma^2_{u-v} + N\sigma^2_{x-y} - 2Nr_{ux} + 2Nr_{uy} + 2Nr_{vx} - 2Nr_{vy}) \\ (3) \qquad &= \frac{N}{4}[2(1 - r_{uv}) + 2(1 - r_{xy}) - 2r_{ux} + 2r_{uy} + 2r_{vx} - 2r_{vy}] \\ &= \frac{N}{2}(2 - r_{uv} - r_{xy} - r_{ux} + r_{uy} + r_{vx} - r_{vy}), \end{aligned}$$

as the value of the $I \times T \times M$ interaction sum of squares.

Turning next to the $I \times T$ interaction sum of squares, note that by assigning to t the explicit values, 1 and 2, the sum of squares may be written as

$$2 \sum_i (\bar{X}_{i1.} - \bar{X}_{i..})^2 + 2 \sum_i (\bar{X}_{i2.} - \bar{X}_{i..})^2,$$

which by a simple procedure ([2], p. 246) reduces to

$$\sum_i (\bar{X}_{i1.} - \bar{X}_{i2.})^2, \quad \text{or} \quad \sum (w - z)^2$$

in simplified notation. This becomes

$$\sum w^2 + \sum z^2 - 2 \sum wz,$$

and since $w = (u + v)/2$ and $z = (x + y)/2$,

$$\sum\left(\frac{u+v}{2}\right)^2 + \sum\left(\frac{x+y}{2}\right)^2 - 2\sum\left(\frac{u+v}{2}\right)\left(\frac{x+y}{2}\right)$$
$$= \tfrac{1}{4}(\sum u^2 + \sum v^2 + 2\sum uv + \sum x^2 + \sum y^2 + 2\sum xy - 2\sum ux - 2\sum uy - 2\sum vx - 2\sum vy).$$

Each term in the foregoing involves either N times a variance (of unity) or N times a correlation coefficient; hence

$$\text{(4)} \qquad \frac{N}{2}(2 + r_{uv} + r_{xy} - r_{ux} - r_{uy} - r_{vx} - r_{vy})$$

is the sum of squares for the $I \times T$ interaction.

The six r's called for in (3) and (4) can, of course, be computed without transforming to standard scores. These r's are required in Yule's correction for attenuation formula ([1], pp. 209-210).

The value of r_{wz}, which ordinarily will be needed for descriptive purposes, can readily be obtained in terms of these six r's by substituting in the appropriate formula for the correlation of sums (or averages):

$$\text{(5)} \qquad r_{wz} = \frac{r_{ux} + r_{uy} + r_{vx} + r_{vy}}{\sqrt{2 + 2r_{uv}}\ \sqrt{2 + 2r_{xy}}}.$$

If the parallel halves of test 1 (also test 2) yield strictly comparable raw scores, i.e., equal sigmas, (5) will give a value exactly equivalent to the correlation between the raw scores of the two tests.

When (3) and (4) are divided by their df's, the two resulting variance estimates lead to

$$\text{(6)} \qquad F_1 = \frac{2 + r_{uv} + r_{xy} - r_{ux} - r_{uy} - r_{vx} - r_{vy}}{2 - r_{uv} - r_{xy} - r_{ux} + r_{uy} + r_{vx} - r_{vy}},$$

with $n_1 = n_2 = N - 1$ as the df's.

Since no assumptions were made in arriving at (6), the only assumptions to be met are those which underlie the analysis of variance technique. For the given situation the assumptions are that the errors of measurement entering into the X_{itm} scores are independently and normally distributed with equal variance. Inequality of variances will occur when the two tests differ as to reliability, hence a valid interpretation of F_1 holds only when the reliabilities are the same (or nearly so). These same assumptions hold for F_2 given below.

A little consideration of the $I \times M$ term in Table 2 leads to the a priori assumption that it will be zero within chance limits, hence its sum of squares may be combined with that for the $I \times T \times M$ term. Note that this pooled sum of squares is nothing more than the residual after taking out the variation

due to individuals and that due to $I \times T$ interaction. This residual sum of squares may be written as

$$\sum_i \sum_t \sum_m (X_{itm} - \bar{X}_{it.})^2.$$

When t is given the explicit values, 1 and 2, and m the values, a and b, this sum of squares may be expressed, in simplified notation, as

$$\sum (u - w)^2 + \sum (v - w)^2 + \sum (x - z)^2 + \sum (y - z)^2.$$

Replacing w by $(u + v)/2$ and z by $(x + y)/2$, this can be simplified to

$$\tfrac{1}{2} \sum (u - v)^2 + \tfrac{1}{2} \sum (x - y)^2,$$

which, since the data are in standard scores, becomes

$$N(1 - r_{uv}) + N(1 - r_{xy}),$$

or

$$N(2 - r_{uv} - r_{xy})$$

as the residual sum of squares. The df for this residual is $2N - 2$, hence

$$F_2 = \frac{2 + r_{uv} + r_{xy} - r_{ux} - r_{uy} - r_{vx} - r_{vy}}{2 - r_{uv} - r_{xy}}, \tag{7}$$

with $n_1 = N - 1$ and $n_2 = 2N - 2$ as the df's.

It is of interest to see what happens to F_2 (and indirectly to F_1) under certain conditions. When measures u and v are strictly parallel and measures x and y are also strictly parallel, it will be seen that the last four r's in the denominator of F_1 will be identical, thus drop out, and leave a denominator equal to that of F_2. By utilizing (5), which holds exactly when the measures are strictly parallel, the last four r's in the numerators of (6) and (7) can be replaced by

$$r_{wz}\sqrt{(2 + 2r_{uv})(2 + 2r_{xy})}.$$

This leads to

$$F_3 = \frac{2 + r_{uv} + r_{xy} - r_{wz}\sqrt{(2 + 2r_{uv})(2 + 2r_{xy})}}{2 - r_{uv} - r_{xy}}. \tag{8}$$

When the two tests are equally reliable, and $r_{uv} = r_{xy} = r$,

$$F_4 = \frac{2 + 2r - r_{wz}(2 + 2r)}{2 - 2r},$$

or

$$F_4 = \frac{(1 + r)(1 - r_{wz})}{1 - r}. \tag{9}$$

The r in (9) is the reliability for scores based on half a test. By the Brown-Spearman formula the reliability, r', of total scores would be

$$r' = 2r/(1 + r),$$

from which

$$r = r'/(2 - r').$$

Substituting in (9)

$$F_4 = \frac{[1 + r'/(2 - r')](1 - r_{wz})}{[1 - r'/(2 - r')]},$$

which simplifies to

$$F_4 = \frac{1 - r_{wz}}{1 - r'}. \tag{10}$$

Now, finally, if the w and z sets of scores are thought of as having been transformed to standard score form, the last expression becomes

$$F_5 = \frac{1 - r_{wz}}{\sigma_e^2}, \tag{11}$$

in which σ_e^2 is the error of measurement variance common to the two tests when both the w and the z scores have been transformed to standard scores and it is assumed that the two tests have equal reliabilities.

It is not argued that sample values of F_3, F_4, and F_5 will follow the F distribution exactly. From (11) it is seen that F_5, and by inference F_1 and F_2, is a function of the extent to which an observed correlation deviates from unity *relative* to error of measurement variance.

In applications the F_1 given by (6) or the F_2 of (7), or their equivalents in terms of variance estimates from Table 2, would be used to determine whether the correlation between two variables was farther below unity than attributable to measurement error. Since F_2 involves a larger df for the error term, it is to be preferred over F_1.

Addenda

After the foregoing was submitted for publication, Lord [3] presented a mathematically elegant but computationally complicated approach to the same problem. His likelihood ratio test and the above proposed F test both assume independence of measurement errors. His assumption of equality of population variances for parallel measures is herein obviated by the transformation to standard scores. His assumption of the equality of certain population covariances, set forth in his equation (2), is equivalent to the a priori assumption of no $I \times M$ interaction, which, however, requires that certain population correlations, instead of covariances, be equal. The F method assumes that the two tests are equally reliable (in the population)

whereas the likelihood ratio test is not thus restricted. Lord's is a large sample significance test whereas the F test may be used for either large or small samples.

Until such time as the mathematical interrelation, if any, of the two solutions has been ascertained, it is of interest to see whether or not the F test yields levels of significance similar to those of the likelihood ratio test when applied to Lord's illustrative examples. For this purpose the probabilities for his normal deviates have been determined to two figures beyond the cipher, and a curve relating p to F has been used for graphical interpolation in order to specify the p's associated with the observed F's as computed by formula (7) above. For the seven examples in Lord's Table 1, each of which involves equal observed reliabilities, we have the following levels of significance (Lord's p's given first in each pair): .024, .023; .0028, .0025; .17, .17; .062, .061; .024, .023; .0062, .0080; .0029, .0025. Admittedly, the graphic determinations of the fourth decimal figures are rough approximations.

Lord also applied his method to data on 649 cases for which the observed reliabilities were different, .669 and .757. Since the F for these data goes far beyond available tabled values, it was transformed to a normal deviate by an approximation method given in Wallis and Roberts ([5], p. 458). This yielded a normal deviate of 6.19, compared to the 5.94 obtained by the likelihood ratio test. For these same data treated as though based on $N = 101$, Lord finds significance at the .010 level whereas the F test leads to .007 by both graphical interpolation and the Wallis-Roberts transformation.

Such close agreement in results as shown in these nine examples may not occur when the equal reliability assumption does not hold, but since the study by Norton, as reported in Lindquist [2], shows that marked heterogeneity of variances (25, 100, and 225) has little effect on F tests (i.e., a p of .01 is near the .02 level and a p of .05 is near the .07 level), it can be presumed that such differences in reliabilities as are apt to be encountered in practice will not seriously disrupt the F test proposed above. Reliabilities of .80 and .95, .60 and .90, .40 and .85, and even .20 and .80, each set of which leads to error variances differing by a factor of four, would not be nearly as extreme as the differences in the Norton study.

REFERENCES

[1] Kelley, T. L. *Statistical method.* New York: Macmillan, 1924.
[2] Lindquist, E. F. *Design and analysis of experiments.* Boston: Houghton Mifflin, 1953.
[3] Lord, F. M. A significance test for the hypothesis that two variables measure the same trait except for errors of measurement. *Psychometrika,* 1957, 22, 207-220.
[4] McNemar, Q. *Psychological statistics.* New York: Wiley, 1949.
[5] Wallis, W. A. and Roberts, H. V. *Statistics.* Glencoe, Ill.: The Free Press, 1956.

Manuscript received 8/20/52

Revised manuscript received 11/9/57

PSYCHOMETRIKA—VOL. 23, NO. 3
SEPTEMBER, 1958

THE KUDER-RICHARDSON FORMULA (21) AS A SPLIT-HALF COEFFICIENT, AND SOME REMARKS ON ITS BASIC ASSUMPTION

SAMUEL B. LYERLY
WASHINGTON, D. C.

Case IV of the Kuder-Richardson series, their formula (21), is derived as a generalized split-half Spearman-Brown coefficient. The basic assumption employed is shown to be sufficient to justify the various assumptions used in derivations by other authors. Some of the implications of this assumption are discussed.

Cronbach [1] showed that the Case III reliability coefficient of Kuder and Richardson [2] is the mean of all possible split-half Spearman-Brown coefficients. This demonstration was useful to students of test theory in bringing the Kuder-Richardson formulation closer into line with the familiar split-half concept. It is the purpose of this paper to present a development of the Kuder-Richardson formula (21) from the split-half point of view and to comment on the basic assumption underlying this useful coefficient.

A test of n items (n assumed to be an even number for convenience) may be split into halves in $n!/[2(n/2)!^2]$ possible ways; with random splitting each possibility may be regarded as equally likely. Suppose all of these splits are made for a given individual and all of the possible pairs of half tests for him are scored. From the well-known theorem concerning samples drawn with replacement from a finite population, the variance of either half-score distribution is

$$\text{(1)} \qquad \text{var} = \frac{X_i(n - X_i)}{4(n - 1)}.$$

Since the correlation between all pairs of half scores (or any sample of them) is -1, their covariance is

$$\text{(2)} \qquad \text{cov} = -\frac{X_i(n - X_i)}{4(n - 1)}.$$

Since the pair of scores resulting from one split is as likely as the pair resulting from any other split, there is no reason for preferring one pair rather than another, so consider the entire set and record them all in a bivariate table or plot them all on a scatter diagram. Now repeat this procedure for each of a sample of N individuals and compute the product moment correlation coefficient over all N sets of pairs of half scores.

When a number of equally numerous distributions are merged, the variance of the resulting total distribution is equal to the average of the

variances of the separate distributions plus the variance of their means. Using this identity and (1),

$$\text{var}_T = \frac{1}{N}\left[\frac{\sum X(n-X)}{4(n-1)} + \frac{\sum X^2}{4} - \frac{(\sum X)^2}{4N}\right]. \tag{3}$$

(All summations in this and following equations are over the N subjects.)

Similarly, when equally numerous bivariate distributions are combined, the total covariance is equal to the average of the separate covariances plus the covariance of the means. In our case each separate bivariate distribution (pairs of half scores for an individual subject) is symmetric about a mean of $(X_i/2,\ X_i/2)$ and the correlation over all N sets of means must be $+\ 1$. Their covariance is therefore $\sum X^2/4N - (\sum X)^2/4N^2$, and thus

$$\text{cov}_T = \frac{1}{N}\left[\frac{\sum X^2}{4} - \frac{(\sum X)^2}{4N} - \frac{\sum X(n-X)}{4(n-1)}\right]. \tag{4}$$

Since (3) is the variance of either set of half scores, the correlation coefficient is simply (4) divided by (3). Performing this division gives

$$r' = \frac{\sum X^2 - (\sum X)^2/N - \sum X(n-X)/(n-1)}{\sum X^2 - (\sum X)^2/N + \sum X(n-X)/(n-1)}$$

$$= \frac{Ns_x^2 + (\sum X^2 - n\sum X)/(n-1)}{Ns_x^2 - (\sum X^2 - n\sum X)/(n-1)},$$

where s_x^2 is the variance of the *total* test scores over the N subjects.

Applying the Spearman-Brown formula, $r = 2r'/(1+r')$, and simplifying,

$$r = \frac{s_x^2 + (\sum X^2 - n\sum X)/N(n-1)}{s_x^2}$$

$$= 1 - \frac{n\sum X - \sum X^2}{N(n-1)s_x^2}, \tag{5}$$

which is the Kuder-Richardson formula (21). Replacing $\sum X$ by NM_x and $\sum X^2$ by $Ns_x^2 + NM_x^2$, where M_x is the mean of the scores, (5) becomes

$$r = 1 - \frac{nM_x - s_x^2 - M_x^2}{(n-1)s_x^2},$$

which simplifies to

$$r = \frac{n}{n-1}\left[1 - \frac{M_x - M_x^2/n}{s_x^2}\right]. \tag{6}$$

Equation (6) is one of the familiar formulas for computing K-R (21) using the mean and variance of the scores and the number of items in the test.

A number of derivations of K-R (21) have appeared in the literature, each using somewhat different sets of stated assumptions. A specific assump-

tion used in this paper is implied in the procedure described above in which all the possible pairs of half scores for all individuals in the sample were used to compute the reliability coefficient. It is assumed not only that a pair of scores resulting from one split is as likely to occur as a pair resulting from another split for a given subject, but that each such pair can occur *in combination with any possible pair for any other subject in the sample.* This suggests one way of stating the difference between the Kuder-Richardson formulas (20) and (21). From the split-half point of view, K-R (20) is the expected value of the coefficient when the *same* split is made for every subject. Making a *different random* split for each subject yields a coefficient which converges toward K-R (21).

The above discussion has been couched in split-half terminology. Other writers on K-R (21) have stated their assumptions in terms of their own approaches to the reliability concept. One basic assumption is required; it can be shown that the assumptions in the various derivations, including Kuder and Richardson's original presentation, can be deduced from it.

The basic assumption for K-R (21) can be stated thus: *In a test of n items which are scored* 0 *or* 1 *for failure or success, respectively, each attempt at an item by a subject is an independent trial in the Bernoulli sense.* A Bernoulli trial is an event which has two possible outcomes, which may be called success or failure. If a number of independent trials are performed, they are indistinguishable so far as probability of success is concerned. Tossing coins, throwing dice, and drawing balls from urns are classic examples.

For mental tests, this assumption implies that individual items have no separate identities, even though each may be worded differently or may ask a question different from that asked by any other item. Furthermore, item equivalence from person to person does not exist. The same item, worded identically, may appear on everyone's test booklet, but item 1 for Subject A is regarded as being no more like item 1 for Subject B than it is like item 2 for Subject B. The two sets of items are, in effect, different samples of items from a common pool or population. Since items do not have individual identities, any method of splitting the test into halves is necessarily random from subject to subject, and the assumption employed in the above split-half derivation follows. For the same reason, any item statistic (difficulty, variance, item-test correlation, etc.) can be computed only by selecting responses from subject to subject at random. These considerations dictate the "equal difficulty, equal correlations, matrix of unit rank" assumptions of Kuder and Richardson. Lord's derivation [3], from the point of view of parallel tests constructed by random sampling from an item pool, implicitly uses this assumption. The K-R (20) formula, on the other hand, "recognizes" items and makes use use of item-subject interaction, which is assumed not to exist in the K-R (21) formulation.

REFERENCES

[1] Cronbach, L. J. Coefficient alpha and the internal structure of tests. *Psychometrika*, 1951, **16**, 297–334.

[2] Kuder, G. F. and Richardson, M. W. The theory of the estimation of test reliability. *Psychometrika*, 1937, **2**, 151–160.

[3] Lord, F. M. Sampling fluctuations resulting from the sampling of test items. *Psychometrika*, 1955, **20**, 1–22.

Manuscript received 7/30/57
Revised manuscript received 1/14/58

PSYCHOMETRIKA—VOL. 23, NO. 3
SEPTEMBER, 1958

THE AVERAGE SPEARMAN RANK CRITERION CORRELATION WHEN TIES ARE PRESENT

EDWARD E. CURETON
UNIVERSITY OF TENNESSEE

This note presents the average Spearman rank correlation between m independent rankings and an untied criterion ranking, corrected for ties in any or all of the independent rankings.

Lyerly [2] has considered the average Spearman rank correlation between m experimentally independent rankings of n persons or objects and a criterion ranking, with no ties in any of the rankings, and has shown that its distribution approaches normality rapidly as m and n increase.

Let y represent a criterion rank, x a rank in any one of the independent rankings, and

$$X = \sum^{m} x. \tag{1}$$

Then X is the sum of the independent rankings of one person or object, i.e., the sum of the independent rankings for fixed y. With this notation, Lyerly's formula may be written,

$$\rho_{av} = 1 - \frac{4n + 2}{n - 1} + \frac{12 \sum^{n} Xy}{m(n^3 - n)}. \tag{2}$$

The total sum of squares of differences between a y and an x, as given by Lyerly, is

$$\sum^{m} \sum^{n} d^2 = \frac{2mn(n + 1)(2n + 1)}{6} - 2 \sum^{n} Xy. \tag{3}$$

Dividing by m to obtain the average and then substituting this average for $\sum d^2$ in the usual formula for the Spearman rank correlation $[\rho = 1 - 6 \sum d^2/(n^3 - n)]$, the result reduces to (2).

Now assume that ties may be present in any or all of the independent rankings. When any one such independent ranking is correlated with the criterion ranking, Kendall ([1], p. 31) shows that $\sum d^2$ must be increased by $\sum (t^3 - t)/12$, where t is the number of persons or objects having equal ranks in each tied set, and the summation is over all sets of ties in the independent ranking. Let

$$T = \sum (t^3 - t), \tag{4}$$

the summation now being over all sets of tied ranks in all m independent rankings. Then (3) must be increased by adding $T/12$ to the terms on the right. Dividing by m to obtain the average value of $\sum d^2$ corrected for ties,

and substituting the resulting expression for $\sum d^2$ in the formula for the Spearman rank correlation, yields

$$\rho_{av} = 1 - \frac{4n + 2}{n - 1} + \frac{24 \sum Xy - T}{2m(n^3 - n)}, \tag{5}$$

the average Spearman rank criterion correlation corrected for ties.

Note that ρ_{av} as defined by (5) does not have limits $\pm$ 1 when ties are present; the ties in any one independent ranking prevent complete agreement with the criterion ranking even if the order of its ranks is otherwise identical with that of the criterion series. This formula is in fact a generalization of Kendall's ρ_a ([1], p. 29). The criterion ranking, moreover, may not contain ties; the problem of the correlation between an independent ranking and a criterion ranking which includes true ties has never to the writer's knowledge been investigated.

Kendall shows ([1], p. 64, footnote) that the standard error of ρ in the null case is the same when ties are present as when they are not. Hence the critical ratio,

$$CR = \rho_{av} \sqrt{m(n - 1)}, \tag{6}$$

may be taken as a unit-normal deviate in testing the hypothesis that the true average correlation is 0, whether or not ties are present. There is no corresponding theory to cover the non-null case.

The approach of the distribution of ρ_{av} to normality with increasing n and m is fairly rapid in the untied case, and should not be appreciably slower when ties are present except in extreme cases. A conservative recommendation would be to use (6) whenever:

(*a*) $n > 5$ and $m + n > 20$;
(*b*) not more than 2/3 of all the ranks in the independent rankings are tied;
(*c*) every independent ranking has at least three distinct values.

When these conditions are met, no correction for continuity is needed; this is fortunate, as the proper correction would be difficult to determine. When the above conditions are not met, there is no good significance test. The applicability of the t approximation to ρ_{av} has not been demonstrated, and direct computation of the exact probability would be prohibitively time consuming even for small m and n when ties are present.

REFERENCES

[1] Kendall, M. G. *Rank correlation methods.* London: Griffin, 1948.
[2] Lyerly, S. B. The average Spearman rank correlation coefficient. *Psychometrika,* 1952, 17, 421–428.

Manuscript received 11/22/57
Revised manuscript received 1/24/58

PSYCHOMETRIKA—VOL. 23, NO. 3
SEPTEMBER, 1958

NOTE ON "EFFICIENT ESTIMATION AND LOCAL IDENTIFICATION IN LATENT CLASS ANALYSIS"*

RICHARD B. MCHUGH
UNIVERSITY OF MINNESOTA

I am indebted to Dr. Albert Madansky for detecting the following two errata and for suggesting the corrections outlined below.

(1) The information functions were incorrectly given ([1], p. 346). They may be corrected as follows:

Write "numerator of $q_{(c)(i_1 i_2 \cdots i_p)}$" as $\hat{q}_{(c)(i_1 i_2 \cdots i_p)}$ and "numerator of $q_{(i;c)(i_1 i_2 \cdots i_p)}$" as $\hat{q}_{(i;c)(i_1 i_2 \cdots i_p)}$. Then

$$\frac{1}{n} I_{(c)(c')} = \sum_D \frac{[\hat{q}_{(c)(i_1 i_2 \cdots i_p)} - \hat{q}_{(\gamma)(i_1 i_2 \cdots i_p)}][\hat{q}_{(c')(i_1 i_2 \cdots i_p)} - \hat{q}_{(\gamma)(i_1 i_2 \cdots i_p)}]}{g_{i_1 i_2 \cdots i_p}}$$

for $c, c' = 1, 2, \cdots, \gamma - 1$;

$$\frac{1}{n} I_{(i;c)(i';c')} = \sum_{d \varepsilon D} \frac{\hat{q}_{(i;c)(i_1 i_2 \cdots i_p)} \hat{q}_{(i';c')(i_1 i_2 \cdots i_p)} g_{(i;c)} g_{(i';c')}}{g_{i_1 i_2 \cdots i_p}(g_{(i;c)} - 1 + \delta_{id})(g_{(i';c')} - 1 + \delta_{i'd})}$$

for $i, i' = 1, 2, \cdots, p$, $c, c' = 1, 2, \cdots, \gamma$; and

$$\frac{1}{n} I_{(c')(i;c)} = \sum_{d \varepsilon D} \frac{[\hat{q}_{(c')(i_1 i_2 \cdots i_p)} - \hat{q}_{(\gamma)(i_1 i_2 \cdots i_p)}] \hat{q}_{(i;c)(i_1 i_2 \cdots i_p)} g_{(i;c)}}{g_{i_1 i_2 \cdots i_p}(g_{(i;c)} - 1 + \delta_{id})}$$

for $c' = 1, 2, \cdots, \gamma - 1$, $c = 1, 2, \cdots, \gamma$, and $1 = 1, 2, \cdots, p$, where

$$\delta_{id} = \begin{cases} 1 & \text{if question } i \text{ is answered } yes \text{ in the } d\text{th member of } D, \\ 0 & \text{otherwise.} \end{cases}$$

The errors arise in differentiating log L with respect to f_c and $g_{i;c}$. When differentiating log L with respect to f_c , $c = 1, \cdots, \gamma - 1$, one must differentiate $f_\gamma = 1 - f_1 - f_2 - \cdots - f_{\gamma-1}$ with respect to f_c as well; when differentiating log L with respect to $g_{i;c}$, one must evaluate

$$\frac{\partial\, [f_c \prod_{i \varepsilon P_d} g_{(i;c)} \prod_{i' \varepsilon \bar{P}_d} (1 - g_{(i';c)})]}{\partial g_{(i;c)}}$$

for each d, where P_d is the set of questions answered "positively" for the dth member of D. After taking these points into account in determining S_c and $S_{i;c}$, the information functions as stated above are easily obtained.

**Psychometrika*, 1956, 21, 331–347.

(2) The statement was also made ([1], p. 337) that if an estimator of the structural parameter θ is consistent, then θ must be locally identifiable. Actually, if the estimator is consistent an even stronger statement can be made, namely that θ is identifiable (see [2], p. 376). However, since the proposed estimator of the latent parameters is consistent only when certain (local) restrictions and regularity conditions on the parameters are met, the fact that the estimator is consistent only implies that the parameters are locally identifiable and not identifiable.

REFERENCES

[1] McHugh, R. B. Efficient estimation and local identification in latent class analysis. *Psychometrika*, 1956, **21**, 331–347.

[2] Reiersøl, O. Identifiability of a linear relation between variables which are subject to error. *Econometrica*, 1950, **18**, 375–389.

Manuscript received 5/16/58

BOOK REVIEWS

HENRY QUASTLER (Editor). *Information Theory in Psychology: Problems and Methods.* Glencoe, Illinois: The Free Press, 1955. Pp. x + 436.

Several patterns for presenting the proceedings of specialized scientific conferences have developed. Differences in patterns can be exemplified in reports on the one topic with which the book under review is concerned—information theory. One extreme is represented by the book published in 1953 by the Josiah Macy, Jr. Foundation, *Cybernetics, Circular Causal and Feedback Mechnisms in Biological and Social Systems,* which seemed to have been a scarcely retouched publication of the working papers as they were actually presented in the conference, together with a practically verbatim transcript of all the discussion. A report of another conference on information theory held in 1954 was multilithed and published more or less informally by J. C. R. Licklider of M.I.T.; this consisted of a somewhat filtered and digested version of what went on, Licklider using himself as a transducer. A more standard pattern was represented by the University of Pittsburgh's *Current Trends in Information Theory* (1953), presenting simply a series of scientific papers, amply edited and proofread by their authors, who were invited to read them at this conference.

The present volume invites favorable comparison with all of the afore-mentioned. Nearly all the working papers presented at the conference are printed as edited by their authors *after* the conference, and interwoven is a sort of running comment by the editor, giving his own reactions as well as relaying the significant points which came up in the discussion at the conference. Furthermore, the papers are arranged not in the order in which they were presented at the conference but in an order which the editor found after the conference to be more logical and advantageous.

The editor is able to make the reader feel that he is not missing any of the really significant discussion, but he cautions that a number of "really good papers" were not published because their authors could not find time to prepare them for publication.

So much for the format of this varityped, offset-printed volume. What of its content? Even though it can hardly be said to have the appearance or organization of a textbook, the book could well become one of the standard references on information theory in psychology, for it represents the first serious attempt to assay the positive gains in this field and to standardize notation and terminology, instead of merely presenting a somewhat disorganized array of interesting suggestions and speculations, as do some of the other works on information theory.

In this respect, a most important feature is the generalization of information theory beyond the bounds of its most frequent application in communication. A few quotations from the summary composed by the conference members will be of interest. "It is basic to information theory that any event is evaluated against the background of the whole class of events that could have happened. Information theory proposes to measure the effect of operations by which a particular selection is made out of a range of possibilities. . . . In information theory, variation is the indispensable basis of selection, discrimination, communication, specification, and related operations. Traditional statistics is largely concerned with what can be done or said in spite of variation; information theory deals with what can be done because of variation. . . . If two objects or events are related, then they must mutually affect each other's range of possibilities. Thus, relatedness amounts to a mutually selective operation. . . . In this way, information theory describes the strength of coupling between components of a system, parts of a structure, members of an organization, and portions of an ordered sequence in time. . . . The relation most commonly studied is communication."

Seen in this light, information theory may provide a way of measuring more objectively the dynamic relations existing among the parts of Gestalten. In fact there were many hints that the conference members were beginning to be concerned with the study of complex dynamic systems, particularly those represented by the human being as a transducer of input.

Perhaps out of modesty, one of the most valuable papers "Standardized Nomenclature, An Attempt," was placed at the end of the section on "foundations," by the editor, who was also its author. Users of information theory are put on notice, regardless of their personal predilections, that the basic system of notation in information theory will involve measures of H (uncertainly or specificity), T (relatedness or communication), A (partialled relatedness), D (constraint), and C (redundancy), together with associated subscripts, parenthetical terms, and the like.

Along with notation, one would demand computational procedures, and they are all here, even tables of $\log_2 n$ and $-p \log_2 p$ provided by Klemmer from an Air Force Technical Report. McGill gives a concise and elegant statement of his well-known work on the relation between multivariate analysis and information theory. There are reports by Miller, Rogers, Green, and Augenstine of noble attempts to determine the properties of sampling distributions of information functions; while the sampling distributions have not yet been determined for the general case, Miller's finding that the basic informational function, H, is a biased statistic should receive wide attention. Attneave shows, however, that under certain conditions a better estimate of H than that achieved by the Miller-Madow correction can be made from averaged values in a transmission matrix. The section on information measures as such comes to a close with a chapter by Henry Quastler on approximate procedures for estimating information measures.

By far the largest portion of the book is devoted to reports of psychological experiments, or to theoretical papers concerning special applications of information theory. In line with the earlier insistence that "communication" is only one area where information theory can be applied, there is almost a complete absence of attention to language or to any kind of symbolic, conventional communication between persons. The closest one comes to finding anything of this nature, in fact, is in Osgood's paper, which reports work on the dimensionality of facial expression as a vehicle for communicating emotion, and in a paper by Fritz and Grier on air traffic control communications. On the contrary, most of the papers are concerned with "communication," if it can be called that, between human organisms and machines. We find, first, that the activity of the organism itself seems to be "unitized," timewise, over what Stroud calls "moments," almost as if human experience existed in a series of frames like those of a motion picture film.

The very interesting question is raised as to how much information can be contained with each moment or frame and this question is further pursued by Hake in a note on the concept of channel capacity. Soon we are looking at the organism starting to interact with a machine—a machine (or rather, an apparatus) which presents, by means of lights or buzzers, something which is in effect a stochastic process; the task of the organism is to judge the relative frequency of events or to predict at any point of time what event will take place next. Two types of theory, the one cognitive, the other associationistic, have been offered to account for the organism's disturbingly nonstrategic behavior in this situation. Further papers, by Senders and Cohen, Deininger and Fitts, and others, discuss the application of information theory to studying the behavior of people reading dials or responding to various other types of information displays. A particularly interesting paper is that by Bricker, who after a review of reaction time experiments suggests that the amount of information processed is linearly related to the time required for processing.

All in all, this is an excellent series of exercises in quantitative rationalizing of psychology.

JOHN B. CARROLL

Harvard University

William G. Cochran and Gertrude M. Cox. *Experimental Designs.* New York: John Wiley & Sons, Inc. Second Edition 1957. Pp. xiv + 611.

This excellent revision of a well-known and successful textbook is better described as a judicial and timely extension than as a reformulation. *Experimental Designs,* first published in 1950, provided an orderly and much needed practical description of experimental designs important for biological research, particularly agricultural experimentation. Cochran and Cox had sought to prepare a handbook of precedures, and from the arithmetical and practical standpoint their *Experimental Designs* was admirably complete. It was not a beginner's cook book, however, and a working knowledge of the statistics involved in experimental analysis, particularly analysis of variance, was assumed.

The present (second) edition, although written in the framework of the original edition, incorporates numerous substantial improvements and extensions. The authors' presentation of the various currently useful experimental plans involves both the introduction of new approaches and the addition of recent developments in well-established procedures. In this way they have sought to accommodate their second edition to the continuously emerging requirements for experimental designs. Because of changes in the present volume and because of the growth of experimentation in industry, this second edition will find many more applications in industry than the original work. Nevertheless, references to specific industrial problems or processes are relatively few. For the most part the references are from the fields of mathematical statistics, biometry, and agriculture.

This is not a volume in which the typical reader from the behavioral sciences will feel at home. Almost all the problems and many of the plans are alien to the experience of the behavior scientist, and for this reason he may not recognize plans which are applicable to his research interest. He will be tempted, moreover, to adapt or modify his research interests to take advantage of some of the elegance and economy represented by many of the plans included in this volume. This is not unusual. Many of the more elegant designs employed by psychologists are borrowed from other fields, and certainly designs developed in other research areas can not be fairly gauged on the basis of their obvious and immediate relevance to the needs of those who investigate behavior.

Although the present edition is sufficiently concerned with specific procedures and explicit cautions to justify its being described as an extremely practical guide to experimental plans, it is not written at the introductory level. Like its predecessor, this book presupposes both sound statistical training and substantial experience in the experimental study of various problems. From the standpoint of those who have not fully mastered basic statistical analysis, Chapter 3 is particularly and significantly instructive. It provides a useful summary of analysis of variance and applications of some of the assumptions involved in least squares procedures.

In addition to the revisions and extensions of the various chapters from the original edition, the present volume contains two new chapters. Chapter 6A deals with the use of fractional replication in factorial experiments. Chapter 8A includes a section on the use of discontinuous data, characteristic of many formal analytical plans, with the interpretation of various functions as continuous trends. The readers will find many new designs in this edition including new incomplete block designs and new incomplete factorial designs. The scope of the volume is best indicated by the chapter headings: 1, Introduction; 2, Methods for Increasing The Accuracy of Experiments; 3, Notes on the Statistical Analysis of the Results; 4, Completely Randomized, Randomized Block, and Latin Square Designs; 5, Factorial Experiments; 6, Confounding; 6A, Factorial Experiments in Fractional Replication; 7, Factorial Experiments with Main Effects Confounded: Split-plot Designs; 8, Factorial Experiments Confounded in Quasi-Latin Squares; 8A, Some Methods for the Study of Response Surfaces; 9, Incomplete Block Designs; 10, Lattice Designs; 11, Balanced and Partially Balanced Incomplete Block Designs; 12, Lattice Squares; 13, Incomplete

Latin Squares; 14, Analysis of the Results of a Series of Experiments; 15, Random Permutations of 9 and 16 Numbers.

It is possible that some of the designs which are appropriate to sequential experimentation in industry may find application in the studies of the behavior scientist. Certainly the economies promised by the various kinds of confounding plans will continue to challenge the psychological investigator.

This book with its handsome format, most readable type, and well-designed presentation of analyses and plans will continue to be greatly appreciated as a textbook. The fact that the new edition comprises additions, refinements, and improvements without changing the structure or the organization of the original volume should make it very convenient for the teacher to introduce the new text into courses which used the old text. The authors are to be commended for undertaking this revision and for their lucid presentation of an intricate body of knowledge. This book can be of great value both to those who perform experiments and to those who responsibly read experimental reports.

J. R. Wittenborn

Rutgers University